

(43)Date of publication of application : 04.11.1997

(51)Int.Cl.

G06F 17/30

(21)Application number : 08-230012

(71)Applicant : MATSUSHITA ELECTRIC IND CO
LTD

(22)Date of filing : 30.08.1996

(72)Inventor : KANEMICHI TOSHIKI
YOSHIDA HIDEYUKI
WATANABE TAISUKE

(30)Priority

Priority number : 08 31547
07226172

Priority date : 20.02.1996
04.09.1995

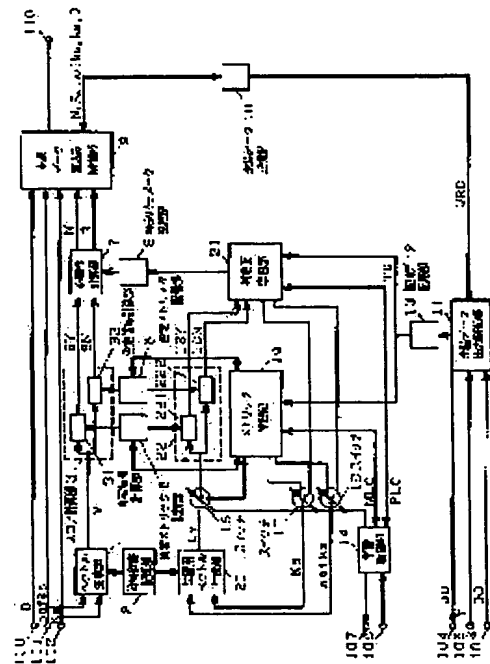
Priority country : JP
JP

(54) INFORMATION FILTERING DEVICE AND INFORMATION FILTERING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide an information filtering device capable of arranging information correspondingly the necessity degree of a user and providing it to the user in an order from the information of high necessity.

SOLUTION: This information filtering device converts plural key words allocated to the information to a vector by a vector generation part 1, calculates a score by using the vector and teacher signals from the user by a score calculation part 3, calculates the necessity and reliability from the score by a necessity calculation part 7 and calculates metric used when the score calculation part 3 calculates the score based on the simple evaluation of the necessity/unnecessity of the information supplied from the user by a metric learning part 19.



LEGAL STATUS

[Date of request for examination]

14.07.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japanese Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-288683

(43) 公開日 平成9年(1997)11月4日

(51) Int.Cl.⁶

G 0 6 F 17/30

識別記号

庁内整理番号

F I

G 0 6 F 15/403

技術表示箇所

3 8 0 E

3 7 0 Z

審査請求 有 請求項の数44 O L (全 35 頁)

(21) 出願番号 特願平8-230012

(22) 出願日 平成8年(1996)8月30日

(31) 優先権主張番号 特願平8-31547

(32) 優先日 平8(1996)2月20日

(33) 優先権主張国 日本 (J P)

(31) 優先権主張番号 特願平7-226172

(32) 優先日 平7(1995)9月4日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 金道 敏樹

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

(72) 発明者 吉田 秀行

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

(72) 発明者 渡辺 泰助

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

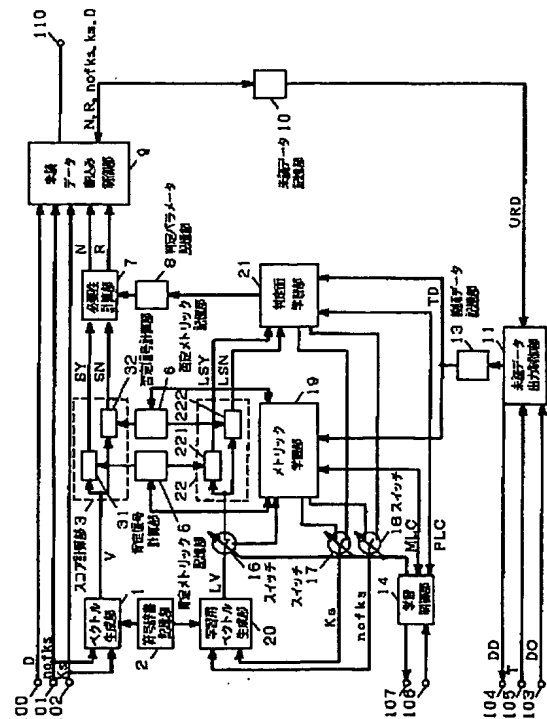
(74) 代理人 弁理士 滝本 智之 (外1名)

(54) 【発明の名称】 情報フィルタ装置及び情報フィルタリング方法

(57) 【要約】

【課題】 情報をユーザーの必要度にしたがって並べ、ユーザーに対して必要性の高い情報から順に提供できる情報フィルタ装置を提供するものである。

【解決手段】 ベクトル生成部1により情報に割り振られた複数のキーワードをベクトルに変換し、スコア計算部3により前記ベクトルと使用者からの教師信号を用いてスコアを計算し、必要性計算部7により前記スコアから必要性和信頼性を計算し、メトリック学習部19によりスコア計算部3がスコアを計算する際に用いるメトリックを使用者から与えられる情報の必要/不要という単純な評価をもとに計算するものである。



【特許請求の範囲】

【請求項1】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すものであって、情報を提示する手段と、前記情報が必要か否かを入力する入力端子を有し、少なくとも前記入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置。

【請求項2】 情報の提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を変換したベクトル信号と、情報が必要か否かという入力とから計算された少なくとも一つのメトリック信号とからスコア信号を計算し、そのスコア信号を利用して情報の提示の順序を決める請求項1記載の情報フィルタ装置。

【請求項3】 メトリック信号は、提示した情報と入力端子から入力される必要か不要かを示す信号とから計算される信号であり、前記入力端子から入力される信号が必要である場合の情報から構成される肯定メトリック信号と、前記入力端子から入力される信号が不要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項2記載の情報フィルタ装置。

【請求項4】 肯定メトリック信号は、入力端子から入力される信号が必要な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項3記載の情報フィルタ装置。

【請求項5】 肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の (ij) 成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要とされた頻度と、前記 i 番目のキーワード信号と前記 j 番目のキーワード信号が同時に含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項3記載の情報フィルタ装置。

【請求項6】 行列の (ij) 成分は、情報が必要であるか不要であるかを示す確率分布と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項5記載の情報フィルタ装置。

【請求項7】 情報の提示の順序づけるために、複数のキーワード信号をベクトル信号に変換する手段と、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ベクトル信号を用いて肯定スコア信号及び否定スコア信号を計算するスコア計算部と、前記肯定スコア信号及び前記否定スコア信号からなる二次元平面の点の分布を必要な情報及び不要な情報に分離する直線の係数である判定パラメータ信号、前記肯定スコア信号並びに前記否定スコア信号から必要性信号及び信頼性信号を計算する必要性計算部を

有し、前記必要性信号の大きさにより情報の提示の順序を決めることを特徴とする請求項1及至6のいずれか記載の情報フィルタ装置。

【請求項8】 判定パラメータ信号は、過去の情報と前記情報が必要か否かという入力との履歴から計算されることを特徴とする請求項7記載の情報フィルタ装置。

【請求項9】 電子ニュース等のように次々と流入してくる情報から必要とする情報を優先的に提示するように、未読の情報を記憶する未読データ記憶部と、前記未読の情報を必要性信号の大きさの順に並べて前記未読データ記憶部に書き込む未読データ書き込み制御部と、前記未読データを順に提示する未読データ出力制御部とを有する請求項1及至8のいずれか記載の情報フィルタ装置。

【請求項10】 情報の必要性を評価するための辞書装置であって、情報が必要か否かを示す信号と前記情報に付けられたひとつまたは複数のキーワード信号とを用いて、情報の必要性を評価するために適するように辞書の内容を更新することを特徴とする適応辞書装置。

【請求項11】 辞書の内容の更新は、必要とされた情報の頻度と、不要とされた情報の頻度と、それぞれのキーワード信号について、前記キーワード信号を含む情報が必要とされた頻度と、前記キーワード信号を含む情報が不要とされた頻度とから計算されるキーワードコスト信号を用いて行われることを特徴とする請求項10記載の適応辞書装置。

【請求項12】 それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項11記載の適応辞書装置。

【請求項13】 それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号が含まれた情報が必要であるか不要であるかを示す確率分布との違いが大きいほど大きな値の信号であり、前記キーワードコスト信号が大きなキーワード信号を残し、小さいキーワード信号を破棄することを特徴とする請求項12記載の適応辞書装置。

【請求項14】 入力端子からの入力が、提示された情報を必要とした回数を示す全肯定回数及び前記情報を不要とした回数を示す全否定回数を記憶する回数記憶部と、キーワード信号を示す文字列を数字に変換する対応表、前記文字列がキーワード信号として含まれた情報を必要とした回数を示す肯定回数及び前記文字列がキーワード信号として含まれた情報を不要とした回数を示す否定回数を記憶した適応辞書記憶部と、提示された情報が必要か否かという入力端子からの入力、前記情報に含まれたキーワード信号、前記全肯定回数、前記全否定回数

及び前記適応辞書記憶部に記憶された信号から前記全肯定回数、前記全否定回数及び前記適応辞書記憶部に記憶された信号を更新する辞書学習部とを有することを特徴とする請求項10及至13のいずれか記載の適応辞書装置。

【請求項15】 請求項10及至14のいずれかに記載された適応辞書装置を有することを特徴とする請求項2及至9のいずれか記載の情報フィルタ装置。

【請求項16】 キーワード信号は、分類コードを含むことを特徴とする請求項2及至9のいずれか、又は15記載の記載の情報フィルタ装置。

【請求項17】 キーワード信号は、分類コードを含むことを特徴とする請求項10及至14のいずれか記載の記載の適応辞書装置。

【請求項18】 請求項17に記載された適応辞書装置を有することを特徴とする請求項16記載の情報フィルタ装置。

【請求項19】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出す方法であって、情報を提示する行程と、前記情報が必要か否かを入力する入力行程を有し、前記入力行程における入力から情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタリング方法。

【請求項20】 情報の提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を変換したベクトル信号と、情報が必要か否かという入力から計算された少なくとも一つのメトリック信号とからスコア信号を計算し、そのスコア信号を利用して情報の提示の順序付けの仕方を変える請求項19記載の情報フィルタリング方法。

【請求項21】 メトリック信号は、入力端子から入力される信号が必要な場合の情報から構成される肯定メトリック信号と、入力端子から入力される信号が不要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項20記載の情報フィルタリング方法。

【請求項22】 肯定メトリック信号は、入力端子から入力される信号が情報が必要な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が情報が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項21記載の情報フィルタリング方法。

【請求項23】 肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の (ij) 成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要とされた頻度と、前記 i 番目のキーワード信号と前記 j 番目のキーワード信号が同時に含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項21記載の情報フィルタリング

方法。

【請求項24】 行列の (ij) 成分は、情報が必要であるか不要であるかを示す確率分布と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項23記載の情報フィルタリング方法。

【請求項25】 情報の必要性を評価するための辞書を構築する方法であって、情報が必要か否かを示す信号と前記情報に付けられたひとつまたは複数のキーワード信号とを用いて、情報の必要性を評価するために適するように辞書の内容を更新することを特徴とする辞書適応方法。

【請求項26】 辞書の内容の更新は、必要とされた情報の頻度と、不要とされた情報の頻度と、それぞれのキーワード信号について、前記キーワード信号を含む情報が必要とされた頻度と、前記キーワード信号を含む情報が不要とされた頻度とから計算されるキーワードコスト信号を用いて行われることを特徴とする請求項25記載の辞書適応方法。

【請求項27】 それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項26記載の辞書適応方法。

【請求項28】 それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であるかを示す確率分布との違いが大きいほど大きな値を信号であり、前記キーワードコスト信号が大きなキーワード信号を残し、小さいキーワード信号を破棄することを特徴とする請求項27記載の辞書適応方法。

【請求項29】 データベースを再構築するデータベース再構築装置であって請求項1及至9、15、16又は18に記載の情報フィルタ装置を用いたことを特徴とするデータベース再構築装置。

【請求項30】 データベースを再構築するデータベース再構築方法であって請求項19及至24に記載の情報フィルタリング方法を用いたことを特徴とするデータベース再構築方法。

【請求項31】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すものであって、情報を提示する手段と、前記情報が必要か否かを入力する入力端子を有し、前記入力端子からの入力に基づいてキーワード検索式を生成することを特徴とするキーワード検索式生成装置。

【請求項32】 情報の提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を変

換したベクトル信号と、情報が必要か否かという入力から少なくとも一つのメトリック信号を計算し、そのメトリック信号を利用してキーワード検索式を生成する請求項31記載のキーワード検索式生成装置。

【請求項33】 メトリック信号は、入力端子から入力される信号が必要な場合の情報から構成される肯定メトリック信号と、入力端子から入力される信号が不要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項32記載のキーワード検索式生成装置。

【請求項34】 肯定メトリック信号は、入力端子から入力される信号が必要な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項33記載のキーワード検索式生成装置。

【請求項35】 肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の(ij)成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、i番目のキーワード信号とj番目のキーワード信号を同時に含む情報が必要とされた頻度と、前記i番目のキーワード信号と前記j番目のキーワード信号が同時に含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項33記載のキーワード検索式生成装置。

【請求項36】 行列の(ij)成分は、情報が必要であるか不要であるかを示す確率分布と、i番目のキーワード信号とj番目のキーワード信号を同時に含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項35記載のキーワード検索式生成装置。

【請求項37】 辞書は、請求項10及至15、または請求項17記載の適応辞書装置によって作成されることを特徴とする請求項31から36いずれかに記載のキーワード検索式生成装置。

【請求項38】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すものであって、情報を提示する手段と、前記情報が必要か否かを入力する入力端子を有し、前記入力端子からの入力に基づいてキーワード検索式を生成することを特徴とするキーワード検索式生成方法。

【請求項39】 情報の提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を交換したベクトル信号と、情報が必要か否かという入力から少なくとも一つのメトリック信号を計算し、そのメトリック信号を利用してキーワード検索式を生成する請求項38記載のキーワード検索式生成方法。

【請求項40】 メトリック信号は、入力端子から入力される信号が必要な場合の情報から構成される肯定メトリック信号と、入力端子から入力される信号が不

要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項39記載のキーワード検索式生成方法。

【請求項41】 肯定メトリック信号は、入力端子から入力される信号が必要な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項40記載のキーワード検索式生成方法。

【請求項42】 肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の(ij)成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、i番目のキーワード信号とj番目のキーワード信号を同時に含む情報が必要とされた頻度と、前記i番目のキーワード信号と前記j番目のキーワード信号を同時に含む情報が不要とされた頻度とから計算されることを特徴とする請求項40記載のキーワード検索式生成方法。

【請求項43】 行列の(ij)成分は、情報が必要であるか不要であるかを示す確率分布と、i番目のキーワード信号とj番目のキーワード信号を同時に含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項42記載のキーワード検索式生成方法。

【請求項44】 辞書は、請求項25及至28のいずれかに記載の辞書適応方法によって作成されるものであることを特徴とする請求項31及至36いずれかに記載のキーワード検索式生成方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、電子または光等を媒体とする記憶装置や情報通信網から必要な情報を取り出し易くする情報フィルタ装置及び情報フィルタリング方法に関するものである。

【0002】

【従来の技術】近年、情報フィルタ装置は、情報通信の社会基盤の進展に伴い、情報通信網の大規模化と通信量の著しい増大に対応する技術として、その実現が強く望まれている。この背景には、今日、個人が処理可能な情報量に対して、個人がアクセスできる情報量が上回るようになっていることがある。このために、大量の情報の中に必要と思う情報が埋没することが、しばしば起こる。

【0003】情報フィルタ装置に関連する従来技術としては、特許検索などに用いられるキーワード論理式をあげることができる。すなわち、数十万から数百万件に及ぶ特許情報をキーワード論理式によりフィルタリングするものである。

【0004】

【発明が解決しようとする課題】しかしながら、キーワード論理式を用いる従来の検索においては、使用者がキ

ワードについての論理式を精度良く設定する必要があるため、使用者がファイリングされているデータ群の癖（例えば、どのような条件の基に、当該データのキーワードが決定されているのか等）やシステムの構造（例えば、キーワードがシソーラス体系のあるシステムであるか否か等）を十分に知り得ていなければ良い検索ができない。このため、初心者には精度の高い情報フィルタリングを行うことができないという課題があった。

【0005】また、情報フィルタリングした結果もキーワードについての論理式に適合するという評価があるだけであり、たまたまキーワードでは合致しているが、内容は求めているものとは異なるケースであったり、あるいは多くの検索結果から使用者にとって必要度の高い情報をその結果から順に取り出すことは容易ではない。

【0006】本発明は上記従来の課題を解決するものであり、初心者にも精度の高い情報フィルタリングができ、かつ使用者にとって必要性の高い情報を取り出し易い情報フィルタ装置を提供することを目的とする。

【0007】

【課題を解決するための手段】この目的を達成するために本発明の情報フィルタ装置は、情報に割り振られた複数のキーワードをベクトルに変換するベクトル生成部と、前記ベクトルと使用者がどんな情報を必要とし不要としたかを表現した行列を用いてスコアを計算するスコア計算部と、前記スコアから必要性和信頼性を計算する必要性計算部と、前記必要性の大きい順に情報を並べ変える未読データ書き込み制御部とからなる情報フィルタリングユニットと、必要性の大きい順に情報を提示し、提示した情報が必要か不要かというユーザーの評価を入力できるインタフェースユニットと、ユーザーの評価と前記複数のキーワードとからスコア計算にもちいる行列を修正する学習ユニットとを含む構成を有している。

【0008】

【発明の実施の形態】この構成によって、複数のキーワードは、距離の定義ができない記号から、使用者の必要度を反映したメトリックを用いて距離を定義できるベクトル表現へと変換され、使用者の必要度を定量化することができ、使用者は必要性の高い情報から順に情報を得ることができるようになる。

【0009】本発明の請求項1に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すものであって、情報を提示する手段と、前記情報が必要か否かを入力する入力端子を有し、少なくとも前記入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置としたものであり、ユーザーからの必要か否かの評価を示す入力を用いて情報の並べ変えを行い情報をユーザーに必要性の高い順に提示するという作用を有する。

【0010】本発明の請求項2に記載の発明は、情報の

提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を変換したベクトル信号と、情報が必要か否かという入力とから計算された少なくとも一つのメトリック信号とからスコア信号を計算し、そのスコア信号を利用して情報の提示の順序を決める請求項1記載の情報フィルタ装置としたものであり、情報の順序づけを情報に付けられた複数のキーワードをベクトルに変換し、そのベクトルとメトリックとからスコアを計算し、そのスコアで情報をユーザーに必要な順に精度高く並べ変えるという作用を持つものである。

【0011】本発明の請求項3に記載の発明は、メトリック信号は、提示した情報と入力端子から入力される必要か不要かを示す信号とから計算される信号であり、前記入力端子から入力される信号が必要である場合の情報から構成される肯定メトリック信号と、前記入力端子から入力される信号が不要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項2記載の情報フィルタ装置としたものであり、メトリックをユーザーが必要とした情報から計算されるものと、不要とした情報から計算されるものの2つを用いることにより、精度の高いスコアを計算でき、情報をユーザーに必要な順に精度高く並べ変えるという作用をもつものである。

【0012】本発明の請求項4に記載の発明は、肯定メトリック信号は、入力端子から入力される信号が必要な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項3記載の情報フィルタ装置としたものであり、これによって簡単な計算でメトリックが計算でき、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0013】本発明の請求項5に記載の発明は、肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の (ij) 成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要とされた頻度と、前記 i 番目のキーワード信号と前記 j 番目のキーワード信号が同時に含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項3記載の情報フィルタ装置としたものであり、精度高くスコア計算ができ、情報をユーザーに必要な順に精度高く並べ変えるという作用を有する。

【0014】本発明の請求項6に記載の発明は、行列の (ij) 成分は、情報が必要であるか不要であるかを示す確率分布と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項5記載の情報フィルタ装置としたものであり、確率分布の違いを評価することによ

り、精度高くスコア計算ができ、情報をユーザーに必要な順に精度高く並べ変えるという作用を有する。

【0015】本発明の請求項7に記載の発明は、情報の提示の順序づけたために、複数のキーワード信号をベクトル信号に変換する手段と、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ベクトル信号を用いて肯定スコア信号及び否定スコア信号を計算するスコア計算部と、前記肯定スコア信号及び前記否定スコア信号からなる二次元平面の点の分布を必要な情報及び不要な情報に分離する直線の係数である判定パラメータ信号、前記肯定スコア信号並びに前記否定スコア信号から必要性信号及び信頼性信号を計算する必要性計算部を有し、前記必要性信号の大きさにより情報の提示の順序を決めることを特徴とする請求項1及至6のいずれか記載の情報フィルタ装置としたものであり、2つのスコア信号を最適に組み合わせることで精度の高い必要性信号を計算でき、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0016】本発明の請求項8に記載の発明は、判定パラメータ信号は、過去の情報と前記情報が必要か否かという入力との履歴から計算されることを特徴とする請求項7記載の情報フィルタ装置2つのスコア信号を最適に組み合わせることで精度の高い必要性信号を計算でき、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0017】本発明の請求項9に記載の発明は、電子ニュース等のように次々と流入してくる情報から必要とする情報を優先的に提示するように、未読の情報を記憶する未読データ記憶部と、前記未読の情報を必要性信号の大きさの順に並べて前記未読データ記憶部に書き込む未読データ書き込み制御部と、前記未読データを順に提示する未読データ出力制御部とを有する請求項1及至8のいずれか記載の情報フィルタ装置であり、ユーザーに必要な情報から優先的に提示するという作用を持つ。

【0018】本発明の請求項10に記載の発明は、情報の必要性を評価するための辞書装置であって、情報が必要か否かを示す信号と前記情報に付けられたひとつまたは複数のキーワード信号とを用いて、情報の必要性を評価するために適するように辞書の内容を更新することを特徴とする適応辞書装置であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0019】本発明の請求項11に記載の発明は、辞書の内容の更新は、必要とされた情報の頻度と、不要とされた情報の頻度と、それぞれのキーワード信号について、前記キーワード信号を含む情報が必要とされた頻度と、前記キーワード信号を含む情報が不要とされた頻度とから計算されるキーワードコスト信号を用いて行われることを特徴とする請求項10記載の適応辞書装置であ

り、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0020】本発明の請求項12に記載の発明は、それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項11記載の適応辞書装置であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0021】本発明の請求項13に記載の発明は、それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号が含まれた情報が必要であるか不要であるかを示す確率分布との違いが大きいほど大きな値の信号であり、前記キーワードコスト信号が大きなキーワード信号を残し、小さいキーワード信号を破棄することを特徴とする請求項12記載の適応辞書装置であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0022】本発明の請求項14に記載の発明は、入力端子からの入力が、提示された情報を必要とした回数を示す全肯定回数及び前記情報を不要とした回数を示す全否定回数を記憶する回数記憶部と、キーワード信号を示す文字列を数字に変換する対応表、前記文字列がキーワード信号として含まれた情報を必要とした回数を示す肯定回数及び前記文字列がキーワード信号として含まれた情報を不要とした回数を示す否定回数を記憶した適応辞書記憶部と、提示された情報が必要か否かという入力端子からの入力、前記情報に含まれたキーワード信号、前記全肯定回数、前記全否定回数及び前記適応辞書記憶部に記憶された信号から前記全肯定回数、前記全否定回数及び前記適応辞書記憶部に記憶された信号を更新する辞書学習部とを有することを特徴とする請求項10及至13のいずれか記載の適応辞書装置。

【0023】本発明の請求項15に記載の発明は、請求項10及至14のいずれかに記載された適応辞書装置を有することを特徴とする請求項2及至9のいずれか記載の情報フィルタ装置であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成され、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0024】本発明の請求項16に記載の発明は、キーワード信号は、分類コードを含むことを特徴とする請求項2及至9のいずれか、又は15記載の記載の情報フィルタ装置であり、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0025】本発明の請求項17に記載の発明は、キーワード信号は、分類コードを含むことを特徴とする請求項10及至14のいずれか記載の記載の適応辞書装置で

あり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0026】本発明の請求項18に記載の発明は、請求項17に記載された適応辞書装置を有することを特徴とする請求項16記載の情報フィルタ装置であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成され、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0027】本発明の請求項19に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出す方法であって、情報を提示する行程と、前記情報が必要か否かを入力する入力行程を有し、前記入力行程における入力から情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタリング方法であり、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0028】本発明の請求項20に記載の発明は、情報の提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を変換したベクトル信号と、情報が必要か否かという入力から計算された少なくとも一つのメトリック信号とからスコア信号を計算し、そのスコア信号を利用して情報の提示の順序付けの仕方を変える請求項19記載の情報フィルタリング方法であり、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0029】本発明の請求項21に記載の発明は、メトリック信号は、入力端子から入力される信号が必要な場合の情報から構成される肯定メトリック信号と、入力端子から入力される信号が不要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項20記載の情報フィルタリング方法であり、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0030】本発明の請求項22に記載の発明は、肯定メトリック信号は、入力端子から入力される信号が情報が必要な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が情報が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項21記載の情報フィルタリング方法であり、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0031】本発明の請求項23に記載の発明は、肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の (ij) 成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要とされた頻度と、前記 i 番目のキーワード信号と前記 j 番目のキーワード信号が同時に含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項21記載の情報フィルタリング方法であり、情報をユー

ザーに必要な順に精度高く並べ変えるという作用を持つ。

【0032】本発明の請求項24に記載の発明は、行列の (ij) 成分は、情報が必要であるか不要であることを示す確率分布と、 i 番目のキーワード信号と j 番目のキーワード信号が同時に含まれた情報が必要であるか不要であることを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項23記載の情報フィルタリング方法であり、情報をユーザーに必要な順に精度高く並べ変えるという作用を持つ。

【0033】本発明の請求項25に記載の発明は、情報の必要性を評価するための辞書を構築する方法であって、情報が必要か否かを示す信号と前記情報に付けられたひとつまたは複数のキーワード信号とを用いて、情報の必要性を評価するために適するように辞書の内容を更新することを特徴とする辞書適応方法であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0034】本発明の請求項26に記載の発明は、辞書の内容の更新は、必要とされた情報の頻度と、不要とされた情報の頻度と、それぞれのキーワード信号について、前記キーワード信号を含む情報が必要とされた頻度と、前記キーワード信号を含む情報が不要とされた頻度とから計算されるキーワードコスト信号を用いて行われることを特徴とする請求項25記載の辞書適応方法であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0035】本発明の請求項27に記載の発明は、それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であることを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であることを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項26記載の辞書適応方法であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0036】本発明の請求項28に記載の発明は、それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であることを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であることを示す確率分布との違いが大きいほど大きな値を信号であり、前記キーワードコスト信号が大きなキーワード信号を残し、小さいキーワード信号を破棄することを特徴とする請求項27記載の辞書適応方法であり、ユーザーの必要とする情報を取り出すために有効な辞書が適応的に構成されるという作用を持つ。

【0037】本発明の請求項29に記載の発明は、データベースを再構築するデータベース再構築装置であって請求項1及至9、15、16又は18に記載の情報フィルタ装置を用いたことを特徴とするデータベース再構築装置であり、ユーザーに必要な情報が取り出し易いとい

う作用を持つ。

【0038】本発明の請求項30に記載の発明は、データベースを再構築するデータベース再構築方法であって請求項19及至24に記載の情報フィルタリング方法を用いたことを特徴とするデータベース再構築方法であり、ユーザーに必要な情報が取り出し易いという作用を持つ。

【0039】本発明の請求項31に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すものであって、情報を提示する手段と、前記情報が必要か否かを入力する入力端子を有し、前記入力端子からの入力に基づいてキーワード検索式を生成することを特徴とするキーワード検索式生成装置であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0040】本発明の請求項32に記載の発明は、情報の提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を変換したベクトル信号と、情報が必要か否かという入力とから少なくとも一つのメトリック信号を計算し、そのメトリック信号を利用してキーワード検索式を生成する請求項31記載のキーワード検索式生成装置であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0041】本発明の請求項33に記載の発明は、メトリック信号は、入力端子から入力される信号が必要な場合の情報から構成される肯定メトリック信号と、入力端子から入力される信号が不要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項32記載のキーワード検索式生成装置であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0042】本発明の請求項34に記載の発明は、肯定メトリック信号は、入力端子から入力される信号が情報が必要な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項33記載のキーワード検索式生成装置であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0043】本発明の請求項35に記載の発明は、肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の (ij) 成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、 i 番目のキーワード信号と j 番目のキーワード信号を同時に含む情報が必要とされた頻度と、前記 i 番目のキーワード信号と前記 j 番目のキーワード信号が同時に含まれた情報が不要とさ

れた頻度とから計算されることを特徴とする請求項33記載のキーワード検索式生成装置であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0044】本発明の請求項36に記載の発明は、行列の (ij) 成分は、情報が必要であるか不要であることを示す確率分布と、 i 番目のキーワード信号と j 番目のキーワード信号を同時に含む情報が必要であるか不要であることを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項35記載のキーワード検索式生成装置であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0045】本発明の請求項37に記載の発明は、辞書は、請求項10及至15、または請求項17記載の適応辞書装置によって作成されることを特徴とする請求項31から36いずれかに記載のキーワード検索式生成装置であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0046】本発明の請求項38に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すものであって、情報を提示する手段と、前記情報が必要か否かを入力する入力端子を有し、前記入力端子からの入力に基づいてキーワード検索式を生成することを特徴とするキーワード検索式生成方法であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0047】本発明の請求項39に記載の発明は、情報の提示の順序づけは、辞書を用いて複数のキーワード信号からなるキーワード群信号を変換したベクトル信号と、情報が必要か否かという入力とから少なくとも一つのメトリック信号を計算し、そのメトリック信号を利用してキーワード検索式を生成する請求項38記載のキーワード検索式生成方法であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0048】本発明の請求項40に記載の発明は、メトリック信号は、入力端子から入力される信号が情報が必要な場合の情報から構成される肯定メトリック信号と、入力端子から入力される信号が不要である場合の情報から構成される否定メトリック信号であることを特徴とする請求項39記載のキーワード検索式生成方法であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0049】本発明の請求項41に記載の発明は、肯定メトリック信号は、入力端子から入力される信号が必要

な場合のベクトル信号の自己相関行列であり、否定メトリック信号は、入力端子から入力される信号が不要である場合のベクトル信号の自己相関行列であることを特徴とする請求項40記載のキーワード検索式生成方法。

【0050】本発明の請求項42に記載の発明は、肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の(ij)成分は、必要とされた情報の頻度と、不要とされた情報の頻度と、i番目のキーワード信号とj番目のキーワード信号を同時に含む情報が必要とされた頻度と、前記i番目のキーワード信号と前記j番目のキーワード信号を同時に含む情報が不要とされた頻度とから計算されることを特徴とする請求項40記載のキーワード検索式生成方法であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0051】本発明の請求項43に記載の発明は、行列の(ij)成分は、情報が必要であるか不要であるかを示す確率分布と、i番目のキーワード信号とj番目のキーワード信号を同時に含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項42記載のキーワード検索式生成方法であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0052】本発明の請求項44に記載の発明は、辞書は、請求項25及至28のいずれかに記載の辞書適応方法によって作成されるものであることを特徴とする請求項31及至36いずれかに記載のキーワード検索式生成方法であり、ユーザーは情報の要不要を入力することにより必要な情報を検索するキーワード検索式を得ることができるという作用を有する。

【0053】以下、本発明の実施の形態について、図1から図19を用いて説明する。

(実施の形態1) 以下、本発明の第一の実施の形態について、図面を参照しながら説明する。図1は本発明の実施の形態1の情報フィルタ装置の構成を示すブロック図であり、図2はその構成と動作を分かりやすくするために機能単位にまとめたブロック図である。

【0054】まず、図2を用いて、本発明の基本概念を説明する。本発明の基本概念の情報フィルタ装置は、ユーザーがどんな「情報」を過去に必要としたかという履歴に関する記録を記憶した複数の記憶部2、5、6、8と、「情報」のフィルタリングを行う情報フィルタリングユニット50と、その情報フィルタリングユニット50により実際にフィルタリングされた未読の「情報」

(ユーザーがまだ読んでいない情報)を蓄積しておく未読データ記憶部10と、ユーザーが当該未読「情報」を可視できるようにしたディスプレイ等のインタフェースユニット51と、ユーザーがどんな「情報」を必要としたかという履歴に関する学習を行う学習ユニット52と

からなる。

【0055】以下、上記構成の動作について説明する。なお、以下の説明では既にユーザーがどんな「情報」を過去に必要としたかという履歴は学習済みのこととして説明する。また、以下に単に「情報」と称するものには、当該「情報」に対応する1つ以上のキーワードが付されているものとする。そのキーワードとは、当該「情報」を構成する各単語の一部あるいは全体であっても良いし、当該「情報」を代表するために特別に付したものであっても良い。

【0056】まず、情報フィルタリングユニット50に新たな「情報」が入力されると、情報フィルタリングユニット50は、記憶部2、5、6、8からユーザーがどのような「情報」を過去に必要としたかという記録を読みだし、前記新たな「情報」の必要性を必要性信号として定量的に評価する。

【0057】次に、その評価された新たな「情報」は、未読データ記憶部10に、必要性信号が大きい順に過去からの未読「情報」を含めて並ぶように前記入力された「情報」を当該順番に書き込む。

【0058】そして、ユーザーが望めば、インタフェースユニット51では、ユーザーに必要性信号の大きい順に前記新たな「情報」を含めた未読「情報」を1つひとつ提示(例えば、ディスプレイに表示)する。

【0059】この際に、ユーザーに提示された前記新たな「情報」を含めた未読「情報」の1つひとつがユーザーにとって必要か不要かを示す教師信号をユーザーがインタフェースユニット51を介して入力することにより、インタフェースユニット51では、当該教師信号を受け取り、当該「情報」とその教師信号を学習ユニット52に送る。なお、このユーザーによる教師信号の入力は、学習ユニット52の学習能力をより高めるために実施するものであり、学習ユニット52の学習能力(ユーザーがどんな「情報」を過去に必要としたかという履歴の学習能力)が既に十分に高ければ行う必要はない。

【0060】次に、学習ユニット52では、前記提示した「情報」とその教師信号を用いて記憶部2、5、6、8の履歴内容を書き換える。

【0061】以上、本発明の情報フィルタ装置は、より高い学習を通じてユーザーに適応し、ユーザーの求める「情報」を優先的に提示することができる。また、当然のことながら、学習を行っていない初期状態では、ユーザーがどんな「情報」を必要としているのか学習ユニット52では分からないので、全ての入力される新たな「情報」をインタフェースユニット51でユーザーが提示を受ける毎に上述したユーザーによる教師信号の入力は必要であるが、随時実施する学習を通じてやがてユーザーに適応し、ユーザーの求める「情報」を優先的に提示することができる。

【0062】なお、ユーザーの求める「情報」を優先的

に提示するとは、より具体的な使用例で述べれば、ある「情報」データベースの母集団Aを特定のキーワードで検索して「情報」の検索集合Bを得たとしても、当該検索集合Bの「情報」の全てがユーザーにとっては必要であるとは限らないし、またユーザーにとっては「情報」の全てが必要であってその必要順位は当然存在することを前提としている。よって、必要から順に不要、あるいはその必要順位に従って、インタフェースユニット51でユーザーに順に提示することを、ユーザーの求める「情報」を優先的に提示することを意味する。

【0063】さて、本発明において重要な点は、いかに必要性信号（或「情報」が必要であったとの教師信号）を計算するかである。

【0064】好ましい実施の形態では、必要性信号は概念的に次のような量として計算される。

【0065】上述べた如く、入力された「情報」にキーワードが添付されている場合を考える。一人のユーザーを考えると、そのユーザーが必要としている「情報」に高い頻度または確率で付いているキーワード集合Aと、不要としている「情報」に高い頻度または確率で付いているキーワード集合Bと、さらにはいずれにもよく付く、または付かないキーワード集合Cとを考えることができる。

【0066】したがって、前記キーワード集合Aに属するキーワードには正の数値を、前記キーワード集合Bに属するキーワードには負の値を、前記キーワード集合Cに属するキーワードには値0をそれぞれ割り振る。

【0067】そして、新たに入力された「情報」について1つ以上のキーワードについてそれぞれが前記キーワード集合A、B、Cのどのキーワードグループに属するかを判定し、前記割り振られた値を積算するように構成する。

【0068】このように構成すれば、前記新たに入力された「情報」に付いていた複数のキーワードを、キーワード集合Aに属するキーワードが数多く含まれた「情報」（ユーザーが必要とする可能性の高い情報）に対しては大きな正の値を示し、キーワード集合Bに属するキーワードが数多く付いている「情報」（ユーザーが不要とする可能性の高い情報）に対しては大きな負の値を示す数値に変換することができる。

【0069】こうして、前記数値を用いてユーザーの必要性に予測することができる。本発明では、提示した「情報」とその「情報」に関するユーザーの必要／不要の評価とからキーワード（キーワード共起を含む）への値の割り振りを自動的に高い精度の高い必要性信号の計算を実現し、精度高く必要性の高い順に「情報」を並べ変えることを実現している。

【0070】そのために、実施の形態1では、「情報」に付けられた複数のキーワードを一つのベクトルに変換し、ユーザーが必要とした場合と不要とした場合につい

て、別々に前記ベクトルの自己相関行列を計算している。ユーザーが必要と答えた「情報」についていたキーワードから作られた自己相関行列MYを用いて、ベクトルVの長さSYを

$$SY = \sum_{i,j} \Sigma M_{ij} \cdot V_i \cdot V_j$$

と計算する。

【0071】なお、以下、必要と答えた「情報」についていたキーワードから作られた自己相関行列MYを「肯定メトリック信号」、不要と答えた情報についていたキーワードから作られた自己相関行列MNを「否定メトリック信号」と呼び、長さSYを肯定信号と呼ぶ。

【0072】この長さSYは、ベクトルVの元となった複数のキーワードの中に、ユーザーが必要とする「情報」によく含まれているキーワードが数多く含まれていれば、長さSYは大きな正の値をとり、そうでない場合には0に近い値をとるから、必要性信号を計算する上で有効である。

【0073】本発明は、以下に図1を用いて詳細説明するように、さらに工夫を重ねて、精度の高い必要性信号の計算を実現している。

【0074】図1を用いて、図2に示した情報フィルタリングユニット50に相当するブロックと、図2に示した学習ユニット52に相当するブロックについて、機能単位を説明しておく。

【0075】まず、情報フィルタリングユニット50に相当するブロックの構成を説明する。

【0076】情報フィルタリングユニット50は、個々の「情報」につけられた複数のキーワード（正確には、分類コードを含む文字列）をベクトルに変換する部分と、ユーザーがどんな「情報」を必要／不要としたという履歴を表現した肯定メトリック信号及び否定メトリック信号を用いてある種のスコアを表す肯定信号と否定信号を計算する部分と、この肯定信号と否定信号とから「情報」の必要性をよく反映する必要性信号を計算する部分と、この必要性信号の大きい順に情報を並べ変える部分からなる。以下、情報フィルタリングユニット50に相当するブロックの構成を、図1に即して説明する。

【0077】図1において、1は「情報」に付けられたキーワードなどの複数の文字列をベクトルに変換するベクトル生成部、2はキーワードなどの複数の文字列をベクトルに変換するための符号辞書信号を記憶した符号辞書記憶部である。この符号辞書記憶部2に記憶された符号辞書信号は、「情報」についているキーワードなどの文字列Wを数字Cに変換する対応表をnofDCK個有するコードブック

【0078】

【数1】

DCK [1] =(W [1], C [1])

・
・
・

DCK [nofDCK] =(W [nofDCK], C [nofDCK])

【0079】であり、ベクトル生成部1はキーワード数信号nofKsとnofKs個のキーワード信号からなるキーワード群信号Ks=(K [1], ..., K [nofKs])とを受けキーワード群信号Ksと前記符号辞書信号DCKを用いてベクトル信号Vに変換する。3はスコア計算部で、ユーザーに提示された「情報」を必要／不要と評価した結果から計算された肯定メトリック信号MY、否定メトリック信号MNを用いて、ベクトル生成部1で変換された2つのベクトル信号Vの長さ、肯定信号SYと否定信号SNに変換する。5は(nofDCK×nofDCK)行列である前記肯定メトリック信号MYを記憶する肯定メトリック記

URD [1] =(N [1], R [1], nofKs [1], Ks [1], D [1])

・
・
・

URD [nofURD] =(N [nofURD], R [nofURD], nofKs [nofURD],
Ks [nofURD], D [nofURD])

【0081】を記憶する未読データ記憶部、13は最大
nofTD個の教師データ信号

【0082】

【数3】

TD [1] =(T [1], TnofK [1], TKs [1])

・
・
・

TD [nofTD] =(T [nofTD], TnofKs [nofTD], TKs [nofTD])

【0083】を記憶する教師データ記憶部である。次に、図2で示したインタフェースユニット51のブロックの構成を説明する。

【0084】図1において、11は制御信号DOを受け未読データ記憶部10から未読データ信号URD [1]を読み出し、表示信号DDを出力し、その表示信号DDがユーザーにとって必要か否かを示す教師信号Tをユーザーから受け、前記教師信号Tと前記未読データ信号URD [1]のキーワード数信号nofKs [1]とキーワード群信号Ks [1]とを所定の手続きに従って教師データ記憶部13に書き込む未読データ出力制御部である。

【0085】次に、図2で示した学習ユニット52に相当するブロックの構成を説明する。学習ユニット52は、ユーザーから入力された教師信号Tを用いて肯定／否定メトリック信号を修正するメトリック学習を行う部分と、肯定／否定信号から必要性信号を計算するためのパラメータ、判定パラメータ信号、を修正する部分からなり、各部分は学習制御部によって制御される。

憶部、6は(nofDCK×nofDCK)行列である前記否定メトリック信号MNを記憶する否定メトリック記憶部である。8は判定パラメータ信号Cを記憶する判定パラメータ記憶部、7は前記肯定信号SYと前記否定信号SNを受け前記判別パラメータ記憶部8から判定パラメータ信号Cを読み出し必要性信号Nと信頼性信号Rを計算する必要性計算部である。9は「情報」の本文である情報データDとキーワード数信号nofKsとキーワード群信号Ksと必要性信号Nと信頼性信号Rとを所定の手続きに従って後述する未読データ記憶部10に書き込む味読データ書き込み制御部、10は前記「情報」の本文である情報データDと前記キーワード数信号nofKsと前記キーワード群信号Ksと前記必要性信号Nと前記信頼性信号Rとからなる最大nofURD個の未読データ

【0080】

【数2】

【0086】図1に示したメトリック学習を行う部分の構成は次のようである。図1において、19は肯定メトリック記憶部5に記憶された肯定メトリック信号MYと前記否定メトリック記憶部6に記憶された否定メトリック信号MNとを修正するメトリック学習部である。このメトリック学習部19は、教師データ記憶部13から前記教師データTDを読み出し、学習ユニット50のベクトル生成部1と同じ機能である学習用ベクトル生成部20で複数のキーワードをベクトルに変換し、自己相関行列を計算することで、肯定／否定メトリック信号を修正する。

【0087】判定パラメータ信号の学習を行う部分の構成は次のようである。図1において、22は学習用肯定信号計算部221と学習用否定信号計算部222とからなる学習用スコア計算部である。この学習用スコア計算部において、221は学習用ベクトル生成部20からの学習用ベクトル信号を受け学習用肯定信号LSYを計算する学習用肯定信号計算部、222は学習用ベクトル生成部20からの学習用ベクトル信号を受け学習用否定信

号LSNを計算する学習用否定信号計算部である。21は学習制御部14からの判定パラメータ学習制御信号PLCを受けて所定の方法で判定パラメータ記憶部8の判定パラメータ信号を書き換える判定面学習部、14は学習開始信号LSを受けてスイッチ16、17、18とメトリック学習部19と学習用ベクトル生成部20と学習用スコア計算部22と学習用否定信号計算部23と判定面学習部21とを制御する学習制御部である。

【0088】以上のように構成された情報フィルタ装置について、各ユニットごとに図面を用いてその動作を説明する。

【0089】情報フィルタ装置の好ましい初期状態の一例は、肯定メトリック信号MYと否定メトリック信号MNとを(nofDCK×nofDCK) 零行列、未読データ記憶部10の未読データURD[i]の全ての必要性信号N[i] (i=1, ..., nofURD) を使用するハードウェアが表現可能な最小の値Vmin、教師データ記憶部13の教師データTD[j]の教師信号T[j]を全て-1とした状態である。

【0090】最初に、情報フィルタリングユニット50の動作を説明する。まず、情報データ入力端子100から情報データDが入力され、キーワード数信号入力端子101から情報データに付けられたキーワードの個数を表すキーワード数信号nofKsが入力され、キーワード信号入力端子102から複数のキーワードであるキーワード群信号Ks=(K[1], K[2], ..., K[nofKs])が入力される。

【0091】ベクトル生成部1によってキーワード群信号Ksは、文字列の集まりからベクトル信号Vへと変換される。この変換によって、キーワード群信号の類似性をベクトルの距離として計算できるようになる。ベクトル生成部1の動作を図3に示すフローチャートを参照しながら説明する。まず、キーワード数信号nofKsとキーワード群信号Ksを受けると(図3ステップ(イ))、内部のベクトル信号V=(V[1], V[2], ..., V[nofDic])を(0, 0, ..., 0)に、キーワードカウンタ信号iを1にセットする(同図ステップ(ロ)、(ハ))。次に、辞書カウンタ信号jを0セットした後辞書カウンタ信号jを1だけ増やす(同図ステップ(ニ))。

【0092】次に、内部にnofDCK個の符号辞書信号DCKを有する辞書記憶部2から辞書カウンタjが指定するキーワードと数字からなる符号辞書信号DCK[j]を読み出し、符号辞書信号DCKの文字列部分W[j]とi番目のキーワード信号K[i]とを比較する(同図ステップ(ホ))。両者が等しくない場合には、辞書カウンタjを1だけ増やす(同図ステップ(ヘ))。両者が一致するか、辞書カウンタjの値が辞書記憶部2に格納された符号辞書信号の個数nofDiCと等しくなるまで図3ステップ(ホ)～(ト)の処理を繰り返す(同図ステッ

プ(ト))。

【0093】キーワード信号K[i]と等しいW[j]が見つかったら、ベクトル信号のj番目の成分V[j]を1にし(同図ステップ(チ))、キーワードカウンタ信号iを1だけ増やす(同図ステップ(リ))。以下、同様の処理をキーワードカウンタ信号iがキーワード数信号nofKsより大きくなるまで実行する(同図ステップ(ヌ))。

【0094】こうして、ベクトル生成部1において、文字列信号からなるキーワード信号の集合体であるキーワード群信号Ksは、0と1でコード化されたnofDCK個の成分を持ったベクトル信号Vに変換される。

【0095】次に、肯定信号計算部31は、キーワード群信号Ksに過去にユーザーの必要とした情報に含まれていたキーワードが数多く含まれる場合に、大きな値となる肯定信号SYを計算する。この目的のために、肯定信号計算部31は、前記ベクトル信号Vを受けて、肯定メトリック記憶部5から肯定メトリック信号MYを読み出し、肯定信号SYを

【0096】

【数4】

$$SY = \sum_{i=0}^{nofDiC-1} \sum_{j=0}^{nofDiC-1} MY[i][j] \cdot V[i] \cdot V[j]$$

【0097】と計算する。否定信号計算部32は、キーワード群信号Ksに過去にユーザーの不要とした情報に含まれていたキーワードが数多く含まれる場合に、大きな値となる否定信号SNを計算する。この目的のために、否定信号計算部32は、否定メトリック記憶部6から否定メトリック信号MNを読み出し、否定信号SNを

【0098】

【数5】

$$SN = \sum_{i=0}^{nofDiC-1} \sum_{j=0}^{nofDiC-1} MN[i][j] \cdot V[i] \cdot V[j]$$

【0099】と計算する。肯定メトリック信号MYと否定メトリック信号MNは、後述するようにキーワード群信号Ksとユーザーの応答に基づいて決められる。本発明では、このように計算された肯定信号SYと否定信号SNを用いて、図9に示したように縦軸に肯定信号SYをとり横軸に否定信号SNをとった2次元空間上の1点に、情報データDを対応させることができる。この2次元空間における情報データDの分布は、ユーザーが必要とするもの(Oで表示)は主に左上部に分布し、ユーザーが不要とするもの(xで表示)は主に右下部に分布するようになる。したがって、図10に示したように適切な係数Cを定めることにより、ユーザーが必要とする情報データDと不要な情報データDとを分離できる。

【0100】さらに、以下に述べるこの係数Cを用いて計算される必要性信号Nは、上述の2次元空間で左上に

ある程、すなわち、必要性の高いと予測される情報データDほど大きな値となる。したがって、必要性信号Nの大きい順に情報データDを並べて提示すれば、ユーザーは必要な情報を効率よく手に入れることができる。必要性信号Nと直交する方向の信頼性信号Rは、大まかにはキーワード群信号Ksに含まれていたキーワードのうちのどれくらいのキーワード信号が辞書に含まれていたかを示す信号である。したがって、この信頼性信号Rの大きさは、情報フィルタが計算した必要性信号Nがどれだけ信頼できるのかを示す。

【0101】次に、必要性計算部7は、前記肯定信号計算部31から出力される前記肯定信号SYと前記否定信号計算部32から出力される前記否定信号SNとを受け、判定パラメータ記憶部8から判定パラメータ信号Cを読み出し、過去必要であった情報についていたキーワードが多数あり、不要であった情報についていたキーワードがほとんどない時に大きな値となる必要性信号Nを $N = SY - C \cdot SN$ と計算し、信頼性信号Rを

$$URD[i+1] = URD[i]$$

と置き換え（同図ステップ（ヘ）～（リ））、その後、i1番目の未読データURD[i1]を

$$\begin{aligned} N[i1] &= N \\ R[i1] &= R \\ \text{nofKs}[i1] &= \text{nofKs} \\ Ks[i1] &= Ks \\ D[i1] &= D \end{aligned}$$

と前記必要性信号N等で置き換える（同図ステップ（ヌ））。この置き換えが終了すると、未読データ部指示端子110から出力する未読データ部指示信号WIを0に戻し（同図ステップ（ル））、処理を終了する。

【0104】次に、未読データURDを読みだし、ユーザーの応答（教師信号T）を付加して教師データ信号TDをつくるインターフェースユニット51について説明する。インターフェースユニット51の動作を図5に示したフローチャートを参照しながら説明する。

【0105】データ読み出し開始信号入力端子103から、データ読み出し開始信号DOが入力される（図5ステップ（イ））。未読データ出力制御部11は、前記未

$$TD[i] = TD[i-1], \quad i=2, \dots, \text{nofTD}$$

と置き換え（同図ステップ（ル））、1番目の教師データTD[1]を前記教師信号Tと前記未読データのキーワード数信号nofKs[1]とキーワード群信号Ks[1]とを用いて $T[1] = T$

$$URD[i] = URD[i+1], \quad i=1, \dots, (\text{nofURD}-1)$$

とし（同図ステップ（ワ）、（カ））、nofURD番目の未読データの必要性信号を

$$N[\text{nofURD}] = (\text{最小値} V_{\min})$$

とする（同図ステップ（ヨ）、（タ）、（レ））。

$$R = C \cdot SY + SN$$

と計算する。

【0102】未読データ書き込み制御部9の動作を、図4に示したフローチャートを参照しながら説明する。まず、それぞれの入力端子から前記情報データDと前記キーワード数信号nofKsと前記キーワード群信号Ksとを受け、必要性計算部7から前記必要性信号Nと前記信頼性信号Rとを受け、未読データ部指示端子110から出力する未読データ処理信号WIを0から1に変える（図4ステップ（イ））。次に、i=1とし（同図ステップ（ロ））、未読データ記憶部10に記憶された未読データURD[i]の必要性信号N[i]（i=1, ..., nofURD）を順次読み出し、前記必要性信号Nと比較し（同図ステップ（ハ））、前記必要性信号Nが未読データURD[i]の必要性信号N[i]より大きくなる（ $N \geq N[i]$ ）最初の未読データの番号i1を検出する（同図ステップ（ニ）（ホ））。

【0103】i1番目以降の未読データを

$$i = i1, \dots, \text{nofURD}$$

読データ記憶部10から1番目の未読データURD

[1]を読み出し（同図ステップ（ロ））、未読データの必要性信号N[1]が最小値Vminより大きい場合には、未読データ信号URD[1]の情報信号D[1]を表示情報信号DDとしてデータ表示端子104に出力し、待機する（同図ステップ（ハ）、（ニ））。未読データの必要性信号N[1]が最小値Vminに等しい場合には、表示情報信号DDを「データなし」としてデータ表示端子104に出力し、待機する（同図ステップ（ホ））。

【0106】ユーザー（図示せず）は、データ表示装置（図示せず）に表示された表示情報信号DDを見て、それが必要な情報である場合には教師信号T=1、必要でない場合には教師信号T=0、処理を終了する場合には教師信号T=-1として、教師信号入力端子105に返す（同図ステップ（ヘ））。教師信号T=-1の場合、処理を終了し、教師信号T≠-1の場合には（同図ステップ（ト））、未読データ出力制御部11は、教師データ記憶部13の（数2）で表わされる教師データを

$$T\text{nofKs}[1] = \text{nofKs}[1]$$

$$TKs[1] = Ks[1]$$

と置き換え（同図ステップ（ル）、（ヲ））、前記未読データ記憶部10の未読データURDを

【0107】次に、学習ユニット52の動作について図6～図8に示したフローチャートを参照しながら説明する。

【0108】図6に学習制御部14の動作の概略を示す

フローチャート示し、詳しく説明する。

【0109】図6において、まず、学習開始信号入力端子106から学習開始信号LSが入力され、学習制御部指示信号出力端子107から出力される学習制御部指示信号LIを0から1に変え(図6ステップ(イ))、処理中を示す。次に、スイッチ16とスイッチ17とスイッチ18とをメトリック学習部19と学習用ベクトル生成部20が接続する様に切り替える(同図ステップ(ロ))。

【0110】次に、図7のステップ(ハ)に対応するメトリック学習部19を動作し(同図ステップ(ハ))、判定面学習部21を動作させた後(同図ステップ(ニ))、LIを0として(同図ステップ(ホ))、処理を終了する。

【0111】次に、メトリック学習部19がユーザーの応答(教師信号T)とキーワード群信号Ksとを用いて、肯定/否定メトリック信号を修正する動作を図7を用いて説明する。

【0112】図7は、メトリック学習部19の動作のフローチャートであり、同図において、前記学習制御部14からメトリック学習制御信号MLCを受けた(図7ステップ(イ))メトリック学習部19は、肯定メトリック記憶部5から肯定メトリック信号MYを、否定メトリック記憶部6から否定メトリック信号MNをそれぞれ読み出す。

【0113】次に、メトリック学習部19は、教師データカウンタcの値を1にする(同図ステップ(ロ))。次に教師データ記憶部13からc番目の教師データ信号TD[c]を読み出し(同図ステップ(ハ))、教師データTD[c]の教師信号T[c]を調べる。前記教師信号T[c]が-1でない場合($T \neq -1$)には(同図ステップ(ニ))、教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群信号TKs[c]とを出力する(同図ステップ(ホ))。前記教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述の情報フィルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力する(同図ステップ(ヘ))。メトリック学習部19は、前記学習用ベクトル信号LVを受け、前記教師データTD[c]の教師信号T[c]が $T=1$ である場合には(同図ステップ(ト))、肯定メトリック信号MYを

$$MY[i][j] = MY[i][j] + LV[i] \cdot LV[j]$$

(ここで、 $i, j = 1 \sim \text{nof DiC}$)と修正する(同図ステップ(チ))。

【0114】この処理により、肯定メトリック信号MYは、ユーザーが必要とした情報データDについていたキーワード信号(複数)に対して大きな値を持つようにな

る。その結果、前述の肯定信号SYが、ユーザーが必要とする情報データDに対して大きくなるようになる。否定メトリック信号MNも以下のように同様の処理がなされる。

【0115】前記教師データTD[c]の教師信号T[c]が $T=0$ である場合には、否定メトリック信号MNを

$$MN[i][j] = MN[i][j] + LV[i] \cdot LV[j]$$

(ここで、 $i, j = 1 \sim \text{nof DiC}$)と修正する(同図ステップ(リ))。

【0116】教師データカウンタの値を

$$c = c + 1$$

と1だけ増やす(同図ステップ(ヌ))。

【0117】以下、メトリック学習部19は、同様の動作を、教師データTD[c]の教師信号T[c]が $T[c] = -1$ になるかまたは $c = \text{nof TD}$ となるまで繰り返す。 $T[c] = -1$ または $c = \text{nof TD}$ となると(同図ステップ(ヲ))、メトリック学習の処理を終了し、メトリック学習制御信号MLCを学習制御部14に送る。

【0118】学習制御部14は、メトリック学習部19からのメトリック学習制御信号MLCを受けて、スイッチ16を学習用ベクトル生成部20とスコア計算部22とが接続するように切り替え、スイッチ17とスイッチ18を学習用ベクトル生成部20と判定面学習部21とが接続するように切り替える。学習制御部14は、判定面学習制御信号PLCを判定面学習部21に送る。

【0119】次に、判定面学習部21について、図8を用いて詳しく説明する。判定面学習部21は、図10に示したように、肯定信号SYと否定信号SNを用いて2次元空間上に表現されたユーザーが必要とする情報データDとユーザーが不要とする情報データDとをもっともよく分離する係数Cを求めるものである。この目的のために、図8に示したフローチャートに従って詳しく説明する。

【0120】まず、前記判定面学習制御信号PLCを受けて(図8ステップ(イ))、教師データカウンタcの値を1にする(同図ステップ(ロ))。教師データ記憶部13からc番目の教師データ信号TD[c]を読み出し(同図ステップ(ハ))、教師データTD[c]の教師信号T[c]を調べる(同図ステップ(ニ))。前記教師信号T[c]が-1でない場合($T \neq -1$)には、教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群信号TKs[c]とを出力する(同図ステップ(ホ))。前記教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述した情報フィルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力す

る。

【0121】学習用スコア計算部22は、前述した情報フィルタリングユニット50のスコア計算部3と同様の動作を行い、学習用肯定信号LSY[c]と学習用否定信号LSN[c]とを出力し、判定面学習部21がそれを受ける(同図ステップ(へ))。前記学習用肯定信号LSY[c]と前記学習用否定信号LSN[c]と教師データTD[c]の教師信号T[c]と判定面学習用信号TC[c]=(T[c], LSN[c], LSY[c])を内部の記憶素子に記憶する(同図ステップ(ト))。そして、教師データカウンタの値を

$$c=c+1$$

と1だけ増やす(同図ステップ(チ))。

【0122】以下、判定面学習部21は、同様の動作を、教師データTD[c]の教師信号T[c]がT[c]=-1になるかまたはc=no of TD+1となるまで繰り返す(同図ステップ(リ))。T[c]=-1またはc=no of TDとなると、学習用肯定信号LSY[c]計算等の処理を終了する。

【0123】次に、判定面学習部21は、内部の記憶素子に記憶された判定面学習用信号TC[c](c=1、...)は、横軸をLSN[c]、縦軸をLSY[c]とし、T[c]=1を○、T[c]=0を×で示すと、図9に示すような分布となる。これらのうち、教師信号T[c]=1であるものと前記教師信号T[c]=0であるものとが、図10に示したように最もよく分離できる判定パラメータCを、山登り法によって計算する(同図ステップ(ヌ))。次に前記判定パラメータCを判定パラメータ記憶部8に書き込み、学習制御部14に判定面学習制御信号PLCを送り(同図ステップ(ル))、処理を終了する。学習制御部14は、判定面学習部21から判定面学習制御信号PLCを受け、学習制御部指示信号を待機中を示す値にし、処理を終了する。

【0124】図10に示したように、上述の2つのメトリック信号を用いてキーワード群信号を肯定信号SYと否定信号SNとで表される2次元空間上で、ユーザーが必要とする情報は主に左上に、不要な情報は右下に分布するようになる。したがって、上記のように適切な係数Cを用いて必要性信号をN=SY-C・SNとすれば、必要性信号は、ユーザーが必要とする情報に対して大きな値をとるようになる。

【0125】なお、判定パラメータCの計算方法として、ここでは、山登り法を採用したが、判定面と学習用必要性信号LNと学習用信頼性信号LRとの距離に基づいて構成されるコスト関数

【0126】

【数6】

$$COST = \sum_c (2 - T[c] - 1)(LSN[c] - C \cdot LSY[c])$$

【0127】を最大にする判定面パラメータCをニュー

トン法、挟み撃ち法などで求める方法であってもよい。

【0128】また、肯定メトリック信号MYと否定メトリック信号MNの学習を忘却の効果を入れた

$$MY[i][j] = \alpha \cdot MY[i][j] + LV[i] \cdot LV[j]$$

$$MN[i][j] = \beta \cdot MN[i][j] + LV[i] \cdot LV[j]$$

を用いてもよい結果が得られる。(ここで、 α と β とは、1より小さい正の数)さらに、文献「情報処理学会技術報告、自然言語処理101-8(1994.5.27)」などに記載された文書からキーワード群信号とキーワード数信号を生成するキーワード生成部を付加する構成をとれば、キーワードが与えられていない情報に対しても適用できる情報フィルタ装置を構成することができる。

【0129】タイトルがつけられた情報については、タイトルを構成する単語をもってキーワードとし、キーワード数信号とキーワード群信号を生成してもよい。

【0130】加えて、キーワード信号は、国際特許分類番号など分類記号を含むようにしても、本発明の構成を変更する必要はなく、よい結果をえることができる。

【0131】また、本発明の実施の形態1では、未読データURDを1つずつ提示する場合について示したが、表示装置(図示せず)の大きさによっては、複数の未読データURDを同時に表示し、使用者が複数表示されたどの未読データに対して応答したのかが正しく情報フィルタ装置に伝えられるような構成を取ることは容易である。

【0132】本発明の情報フィルタの根幹は、図7のフローチャートのト、チ、リに示したように、ユーザーの応答とキーワードとの関係をキーワードの同時出現に注目した肯定メトリック信号MY、否定メトリック信号に反映させ、この2つのメトリック信号を用いてキーワード群信号を肯定信号SYと否定信号SNとに変換することで、キーワードという記号情報を距離の定義された空間に射影したものである。これによって、キーワード群の遠近を距離というアナログ尺度で評価することができるようになる。これを利用することにより、従来の技術では必要か不要かの二者択一的な判定しかできなかった必要性の評価が、ユーザーの必要性の順番に並べるといったことが可能になる。

【0133】本発明の実施の形態1の情報フィルタ装置によれば、ユーザーからの教師信号に基づいた学習によって、ユーザーの必要とする情報に対しては、必要性信号が大きな値を取るようになり、その結果、表示装置等のインターフェースユニットには、ユーザーにとって必要性が高い情報が優先的に表示されるようになる。

【0134】(実施の形態2)以下、本発明の第2の実施の形態について、図面を参照しながら説明する。実施の形態2は、実施の形態1の構成に辞書学習部を付加

し、辞書記憶部2に記憶された符号辞書信号DCKが使用者に適應するように更新し、かつ肯定メトリック信号MYと否定メトリック信号MNを単純な頻度分布に対応するキーワードの自己相関行列から、情報が必要／不要の出現するキーワードの確率分布を考慮したものへと改良したものである。

【0135】図11に本発明の実施の形態2の情報フィルタ装置のブロック結線図を示すが、前述した本発明の実施の形態1の情報フィルタ装置のブロック結線図と異なる構成について詳細に説明する。

【0136】図11において、23は学習制御部14か

$$FDCK[i] = (W[i], C[i], PY[i], PN[i])$$

・
・
・

$$FDCK[nofDCK] = (W[nofDCK], C[nofDCK], PY[nofDCK], PN[nofDCK])$$

【0138】を記憶した適応符号辞書信号記憶部、25は使用者が必要と答えた回数を示す全肯定回数信号NYと不要と答えた回数を示す全否定回数信号NNを記憶する回数記憶部、26は肯定メトリック更新用の1次肯定メトリック信号MY1を記憶する1次肯定メトリック記憶部、27は否定メトリック更新用の1次否定メトリック信号MN1を記憶する1次否定メトリック記憶部、28は前記肯定回数信号と前記否定回数信号と前記1次肯定メトリック信号MY1と前記1次否定メトリック信号MN1とから改良された肯定メトリック信号MYと否定メトリック信号MNを計算してそれぞれを肯定メトリック記憶部5と否定メトリック記憶部6に書き込むKDメトリック学習部である。

【0139】以上のように構成された情報フィルタ装置について、図面を用いて動作を説明する。ただし、動作が実施の形態1と同様の箇所は説明を省略する。

【0140】情報フィルタ装置の好ましい初期状態の一例は、肯定メトリック信号MYと否定メトリック信号MNとを(nofDCK×nofDCK) 零行列、未読データ記憶部10の未読データURD[i]の全ての必要性信号N[i] (i=1, ..., nofURD) を使用するハードウェアが表現可能な最小の値Vmin、教師データ記憶部13の教師データTD[j]の教師信号T[j]を全て-1、適応符号辞書信号の文字列Wを全てブランク、数字Cを符号辞書信号FDCKの上から順に1、2、..., nofDCK、肯定回数PYと否定回数PNを0、適応符号辞書に対応して、符号辞書の文字列も全てブランクとした状態である。

【0141】まず、情報フィルタリングユニット50の動作を説明する。上述の初期状態の場合、実施の形態1に記載した通りの動作を情報フィルタリングユニット50は行い、入力されたキーワード数信号nofKs、キーワ

らの辞書学習信号DLCを受け辞書記憶部2の符号辞書信号DCKを更新する辞書学習部、24は文字列Wと数字Cがキーワード群信号Ksに含まれていたときに使用者が情報データDを必要と解答した回数を示す肯定回数PYと文字列Wがキーワード群信号Ksに含まれていたときに使用者が情報データDが不要と解答した回数を示す否定回数PNとからなる表をnofDCK個有する適応符号辞書信号

【0137】

【数7】

ード群信号Ks、情報データDから必要性信号N、信頼性信号Rとともに0と計算し、未読データ記憶部10に格納する。

【0142】次に、インタフェースユニット51は、実施の形態1と同じ動作を行い、使用者の応答が付いた教師データTDを教師データ記憶部13に送る。

【0143】学習ユニット52の動作は、まず、学習開始信号入力端子106から学習開始信号LSが入力される。学習制御部14は、前記学習開始信号LSを受けて、学習制御部指示信号出力端子107から出力される学習制御部指示信号LIを0から1に変え、処理中を示す。更に、辞書学習信号DLCを辞書学習部23に送る。

【0144】以下、図12に示したフローチャートを参照しながら辞書学習部23の動作を説明する。まず、辞書学習信号DLCを受けて(図12ステップ(イ))、適応符号辞書記憶部24から適応符号辞書FDCKを最大nofDCKtmp個の適応符号信号を記憶できる適応符号信号バッファに読み込み、回数記憶部25から全肯定回数信号NYと全否定回数信号NNとを、1次肯定メトリック記憶部26から1次肯定メトリック信号MY1を、1次否定メトリック信号記憶部27から1次否定メトリック信号MN1を読み出す(同図ステップ(ロ))。次に内部の教師データカウンタcの値を1にし(同図ステップ(ハ))、教師信号記憶部13から教師データTD[c]を読み出し(同図ステップ(ニ))、その教師信号T[c]が-1であるか否かを判定する(同図ステップ(ホ))。

【0145】T[c]≠-1の場合、以下の処理を行う。まず、内部のキーワード数カウンタiの値を1にセットし(同図ステップ(ヘ))、適応符号辞書カウンタjの値を1にセットする(同図ステップ(ト))。次

に、前記文字列W[j]がblankであるかないかを判定し(同図ステップ(チ))、blankである場合には、前記文字列W[j]を前記キーワード信号TK[i]で置き換える(同図ステップ(リ))。blankでない場合には、教師データTD[c]のi番目のキーワード信号TK[i]とj番目の適応符号辞書信号FDCK[j]の文字列W[j]とを比較する(同図ステップ(ヌ))。

【0146】前記文字列W[j]がblankの場合、または、blankでなくかつ前記キーワード信号TK[i]と前記文字列W[j]が一致した場合、T[c]の値に応じて以下の処理を行う。T[c]=1の場合(同図ステップ(ル))、全肯定信号NYに1を加え(同図ステップ(ヲ))、適応符号辞書信号FDCK[j]の肯定回数PY[j]に1を加える(同図ステップ(ワ))。T[c]≠1、これはT[c]=0の場合であるが、全否定信号NNに1を加え(同図ステップ(カ))、適応符号辞書信号FDCK[j]の否定回数PN[j]に1を加える(同図ステップ(コ))。

【0147】前記W[j]がblankでなくかつ前記キーワード信号TK[i]と前記文字列W[j]が一致しない場合、適応符号辞書カウンタjの値を1増やす(同図ステップ(タ))。適応符号辞書カウンタjの値が適応符号辞書信号バッファに記憶できる適応符号信号の数に1を加えた値nofFDCKtmp+1と比較する(同図ステップ(レ))。適応符号辞書カウンタjの値が、nofFDCKtmp+1以下の場合、文字列W[j]がblankかどうかの判定に戻る。

【0148】それ以外の場合は、前記キーワードカウンタiの値を1だけ増やす(同図ステップ(ソ))。

【0149】前記キーワードカウンタiの値が、前記教師データTD[c]のキーワード数信号TnofKSに1を加えた値TnofKS+1と比較して小さい場合(同図ステップ(ツ))、辞書カウンタjを1にセットし、同様の処理

$$NY / (NY + NN) \cdot \log \{ (PY[j]) / (PY[j] + PN[j]) \}$$

$$+ NN / (NY + NN) \cdot \log \{ (PN[j]) / (PY[j] + PN[j]) \}$$

【0154】が考えられる。しかし、これは、そのままでは、本情報フィルタ装置の初期状態など、全肯定回数信号NY、全否定回数信号NN、肯定回数PY[j]、否定回数PN[j]が0のときには、log()の計算ができない、

$$PY[j] + PN[j] \div 1$$

$$KD[j]$$

$$= \tanh \{ (PY[j] + PN[j]) / PC \} \cdot$$

$$\tanh \{ NY / (NY + NN) \}$$

$$\cdot \log \{ (PY[j] + \epsilon) / (PY[j] + PN[j] + 2\epsilon) \}$$

$$+ NN / (NY + NN)$$

$$\cdot \log \{ (PN[j] + \epsilon) / (PY[j] + PN[j] + 2\epsilon) \}$$

を行う。それ以外の場合、教師データカウンタcの値を1だけ増やす(同図ステップ(ネ))。教師データカウンタcの値と教師データ数nofTDに1を加えた値nofTD+1とを比較し(同図ステップ(ナ))、教師データカウンタcの値が小さい場合、次の教師データTD[c]を読み出し同様の処理を行う。

【0150】以上の処理が、全ての教師データTDに対して行われる。次に、辞書学習部23は、各々の適応符号辞書信号FDCK[j]に対し、キーワードコスト信号KDを計算する。このキーワードコスト信号は、文字列W[j]がキーワードとして有効であるか否かを判断するために用いられる量である。

【0151】ところで、使用者の不要な情報データDが出現する確率

$$NN / (NY + NN)$$

と比較して、文字列W[j]が付いている情報データDが使用者にとって不要である場合の確率

$$PN[j] / (PY[j] + PN[j])$$

が大きく異なる場合に、大きくなるようものであれば、文字列W[j]は、情報データDが使用者にとって不要と判定する上で有効である。同様に、使用者の必要な情報データDが出現する確率

$$NY / (NY + NN)$$

と比較して、文字列W[j]が付いている情報データDが使用者にとって必要である場合の確率

$$PY[j] / (PY[j] + PN[j])$$

が大きく異なる場合に、大きくなるようものであれば、文字列W[j]は、情報データDが使用者にとって必要と判定する上で有効である。

【0152】キーワードコスト信号KDは、この性質を反映している量で有ればなんでもよいが、好ましい例の一つとして、カルバックダイバージェンスと呼ばれる

$$【0153】$$

$$【数8】$$

を満たす適応符号辞書信号FDCK[j]のキーワードコスト信号を過大評価する等不適切な場合がある。これを回避する好ましい実施の形態の一つは、キーワードコスト信号を

$$【0155】$$

$$【数9】$$

【0156】とするものである。ここで、 ε は0でのわり算、 $\log 0$ を避けるための小さな正の値を持つパラメータである。パラメータPCは、3程度の値とするとい。

【0157】次に、適応符号辞書信号FDCK[j]の文字列W[j]と肯定回数PY[j]と否定回数PN[j]とをキーワードコスト信号KDの大きい順に並べ替える(同図ステップ(ラ))。このとき、適応符号辞書FDCK[j]の数字C[j]には、最初の並び順が残っている。これを利用して、1次肯定メトリック信号MY1とC[j]から、C[i]、C[j]の値がともに符号辞書DCKの数nofDCKより小さい場合、 $M[i][j] = MY1[C[i]][C[j]]$, $i, j = 1, \dots, \text{nofDCK}$

その他の場合は、 $i = j$ の場合は、 $M[i][i] = PY[C[i]]$, $i = 1, \dots, \text{nofDCK}$

$i \neq j$ の場合は、

$M[i][j] = 0$, $i, j = 1, \dots, \text{nofDCK}$

とした上で、

$MY1[i][j] = M[i][j]$, $i, j = 1, \dots, \text{nofDCK}$

と1次肯定メトリック信号MY1の置き換えを行う。1次否定メトリック信号MN1に対しても、同様の置き換えを行う(同図ステップ(ム))。

【0158】そして、適応符号辞書信号バッファ内の適応符号辞書FDCK[j]の数字C[j]を $C[j] = j$, $j = 1, \dots, \text{nofDCKtmp}$ と置き換える。

【0159】以上の処理を終えると、辞書学習部23は、適応符号辞書バッファ内の適応符号辞書FDCKの上位nofDCK個の文字列W[j]と数字C[j]を辞書記憶部2に書き込み、適応符号辞書バッファ内の適応符号辞書信号FDCK[j]の上位nofDCK個を適応符号辞書記憶部24に書き込み、全肯定回数信号NYと全否定回数信号NNを回数記憶部25に書き込み、1次肯定メトリック信号MY1を1次肯定メトリック信号記憶部26に1次否定メトリック信号MN1を1次否定メトリック信号記憶部27に書き込む(同図ステップ(ウ))。

【0160】最後に、辞書学習信号DCLを学習制御部14に戻して(同図ステップ(ヒ))、処理を終了する。

【0161】次に、前記学習制御部14は、スイッチ16とスイッチ17とスイッチ18とをメトリック学習部19と学習用ベクトル生成部20が接続する様に切り替える。前記学習制御部14は、KDメトリック学習部28にメトリック学習制御信号MLCを送る。

【0162】前記メトリック学習制御信号MLCを受けたKDメトリック学習部28は、まず、1次肯定メトリック記憶部26から1次肯定メトリック信号MY1を、

1次否定メトリック記憶部27から1次否定メトリック信号MN1をそれぞれ読み出す。

【0163】次に、KDメトリック学習部28は、教師データカウンタcの値を1にする。教師データ記憶部13からc番目の教師データ信号TD[c]を読み出し、教師データTD[c]の教師信号T[c]を調べる。前記教師信号T[c]が-1でない場合($T \neq -1$)には、教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群信号TKs[c]とを出力する。前記教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述した実施の形態1の情報フィルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力する。KDメトリック学習部28は、前記学習用ベクトル信号LVを受け、前記教師データTD[c]の教師信号T[c]が $T = 1$ である場合には、1次肯定メトリック信号MY1を $MY1[i][j] = MY1[i][j] + LV[i] \cdot LV[j]$

(ここで、 $i, j = 1 \sim \text{nofDiC}$)と修正する。前記教師データTD[c]の教師信号T[c]が $T = 0$ である場合には、1次否定メトリック信号MN1を $MN1[i][j] = MN1[i][j] + LV[i] \cdot LV[j]$

(ここで、 $i, j = 1 \sim \text{nofDiC}$)と修正する。教師データカウンタの値を

$c = c + 1$

と1だけ増やす。

【0164】以下、KDメトリック学習部28は、同様の動作を、教師データTD[c]の教師信号T[c]が $T[c] = -1$ になるかまたは $c = \text{nofTD}$ となるまで繰り返す。T[c] = -1または $c = \text{nofTD}$ となると、1次肯定メトリック信号MY1と1次否定メトリック信号MN1の学習を終える。

【0165】次に、回数記憶部25から全肯定回数信号NYと全否定回数信号NNを読み出し、1次肯定メトリック信号MY1と1次否定メトリック信号MN1とを用いて、肯定メトリック信号MYを計算する。

【0166】こうして計算される肯定メトリック信号MY、否定メトリック信号MNは、キーワードコスト信号KDと同様、計算される肯定信号SYと否定信号SNが、使用者の不要な情報データDが出現する確率 $NN / (NY + NN)$

と比較して、文字列W[j]が付いている情報データDが使用者にとって不要である場合の確率

$PN[j] / (PY[j] + PN[j])$

が大きく異なる場合に、大きくなるようものであり、使用者の必要な情報データDが出現する確率

$NY / (NY + NN)$

と比較して、文字列W[j]が付いている情報データDが使用者にとって必要である場合の確率
 $PY[j] / (PY[j] + PN[j])$
 が大きく異なる場合に、大きくなるようものであるとい

MY[i][j]

$= NY / (NY + NN)$

$\cdot \log \{ (MY1[i][j] + \epsilon) \cdot (NY + NN) \}$

$/ (NY \cdot (MY1[i][j] + MN1[i][j] + 2\epsilon))$

【0168】と計算し、否定メトリック信号MNを
 【0169】

MN[i][j]

$= NN / (NY + NN)$

$\cdot \log \{ (MN1[i][j] + \epsilon) \cdot (NY + NN) \}$

$/ (NN \cdot (MY1[i][j] + MN1[i][j] + 2\epsilon))$

【0170】と計算する。ここで、 ϵ は0でのわり算、log0を避けるための小さな正の値を持つパラメータである。

【0171】そして、更新された1次肯定メトリック信号MY1を1次肯定メトリック信号記憶部26に、更新された1次否定メトリック信号MN1を1次否定メトリック信号記憶部27に、新たに計算された肯定メトリック信号MYを肯定メトリック記憶部5へ、新たに計算された否定メトリック信号MNを否定メトリック記憶部6へ書き込む。以上で、KDメトリック学習部28は、メトリック学習の処理を終了し、メトリック学習制御信号MLCを学習制御部14に送る。

【0172】学習制御部14は、KDメトリック学習部28からのメトリック学習制御信号MLCを受けて、スイッチ16を学習用ベクトル生成部20とスコア計算部22とが接続するように切り替え、スイッチ17とスイッチ18を学習用ベクトル生成部20と判定面学習部21とが接続するように切り替える。学習制御部14は、判定面学習制御信号PLCを判定面学習部21に送る。

【0173】判定面学習部21の動作は、実施の形態1と全く同じであるので、説明は繰り返さない。

【0174】一度、以上の処理が行われると、辞書記憶部2の符号辞書が空でなくなるので、情報フィルタリングユニット50から出力される必要性信号N、信頼性信号Rは、0でなくなり、使用者の必要性の高い情報データが、未読データ記憶部10の上位に書き込まれるようになる。

【0175】以後、上記処理を繰り返すことにより、使用者が必要とする情報が否かを判定するために有効なキーワードが優先的に辞書記憶部2に記憶されるようになり、小規模な辞書であっても、精度の高い情報フィルタリングが可能となる。

【0176】なお、判定パラメータCの計算方法として、ここでは、山登り法を採用したが、実施の形態1と

った性質を持っていれば、なんでもよい。これを満たす好ましいのは、肯定メトリック信号MYを

【0167】

【数10】

【数11】

同様、判定面と学習用必要性信号LNと学習用信頼性信号LRとの距離に基づいて構成されるコスト関数を最大にする判定面パラメータCをニュートン法、挟み撃ち法などで求める方法であってもよい。さらに、簡便な方法として、

$C = \tan \theta_i$

ここで、

$\theta_i = 0.5 \cdot \pi (i/90) \quad i = 1, \dots, 90$

の中から、 $T[c] = 1$ である情報と $T[c] = 0$ である情報をもっともよく分離できるCを選ぶと言う方法も考えることができる。

【0177】また、1次肯定メトリック信号MY1と1次否定メトリック信号MN1の学習を忘却の効果を入れた

$MY1[i][j] = \alpha \cdot MY1[i][j] + LV[i] \cdot LV[j]$

$MN1[i][j] = \alpha \cdot MN1[i][j] + LV[i] \cdot LV[j]$

を用いてもよい結果が得られる。(ここで、 α は、1より小さい正の数)もしくは、MY1[i][j]またはMN1[i][j]のいずれかが一定値をこえた場合に、

$MY1[i][j] = MY1[i][j] / 2$

$MN1[i][j] = MN1[i][j] / 2$

として、信号のオーバーフローを防ぐように構成することは、実施上好ましい。これは、適応符号辞書信号FDCK[j]の肯定回数PY[j]と否定回数PN[j]、および全肯定回数信号NYと全否定回数NNについても同様である。

【0178】さらに、文献「情報処理学会技術報告、自然言語処理101-8(1994.5.27)」などに記載された文書からキーワード群信号とキーワード数信号を生成するキーワード生成部を付加する構成をとれば、キーワードが与えられていない情報に対しても適用

できる情報フィルタ装置を構成することができる。

【0179】タイトルがつけられた情報については、タイトルを構成する単語をもってキーワードとし、キーワード数信号とキーワード群信号を生成してもよい。

【0180】加えて、キーワード信号は、国際特許分類番号など分類記号を含むようにしても、本発明の構成を変更する必要はなく、よい結果をえることができる。

【0181】また、本実施の形態では、未読データURDを一つずつ提示する場合について示したが、表示装置(図示せず)の大きさによっては複数の未読データURDを同時に表示し、使用者がどの未読データURDについて応答したのかを正しく情報フィルタ装置に伝える構成をとることは容易である。

【0182】以上、本発明の実施の形態2の情報フィルタの根幹は、キーワードの同時出現に注目したメトリックを導入することにより、キーワードという記号情報を距離の定義された空間に射影したことにある。これによって、キーワード群の遠近を距離というアナログ尺度で評価することができるようになる。これを利用することにより、従来の技術では必要か不要かの二者択一的な判定しかできなかった必要性の評価が、ユーザーの必要性の順番に並べるといったことが可能になる。

【0183】本実施の形態による情報フィルタによれば、ユーザーからの教師信号に基づいた学習によって、ユーザーの必要とする情報に対しては、必要性信号が大きな値を取るようになり、その結果、表示装置等には、ユーザーにとって必要性が高い情報が優先的に表示されるようになる。

【0184】(実施の形態3)以下、本発明の第3の実施の形態について、図面を参照しながら説明する。実施の形態3は、本発明の実施の形態1の情報フィルタ装置の構成にデータベース再構築制御部、データベース読み出し部及び適応データベース書き込み部等を付加し、情報フィルタ装置データベース再構築装置としたもので、実施の形態1のインタフェースユニット51、学習ユニット52及び情報フィルタリングユニットの機能を用いて使用者にとって必要な順にデータが並んだ使いやすい適応データベースを提供するものである。

【0185】図13に本発明の実施の形態3のデータベース再構築装置ブロック結線図を示し、以下に説明する。

【0186】図13において、60はデータベース記憶部、61はデータベース記憶部60からデータを読み出し情報フィルタ装置に適した形にデータを整形して出力するデータベース読み出し部、62はデータベースの再構築を制御するデータベース再構築制御部、63はスイッチ、65は適応データベース記憶部、64は情報フィルタリングユニットからの信号を一時的に保持し最終結果を適応データ記憶部に書き込む適応データベース書き込み部、200はデータベース再構築制御部62を制御

する制御信号入力端子、201は学習データ数信号を入力する学習数信号入力端子である。その他のものは、実施の形態1に記載した情報フィルタ装置と同一の構成であるので省略する。

【0187】以上のように構成されたデータベース再構築装置の動作について説明する。まず、制御信号入力端子200からデータベース再構築開始を示す制御信号CDB=1

が入力され、学習数信号入力端子201から情報フィルタ装置の学習回数を示す学習数信号LNが入力される。データベース再構築制御部62は、データベース再構築制御指示信号出力端子210から出力されるデータベース再構築制御指示信号IRDを0から1に変え、処理中であることを示す。データベース再構築制御部62は、スイッチ63を情報フィルタリングユニット50と未読データ記憶部10とを接続するように切り替える。データベース再構築制御部62は、制御信号CDB(=1)を受けて、データベース読み出し部61にデータベース記憶部60に記憶されたデータの数を聞く。データベース読み出し部60は、データベース記憶部60に記憶されたデータの数を数えその結果をデータ数信号nofDとして、データベース再構築制御部62に送る。データベース再構築制御部62は、適応データベース書き込み制御部64内の未読データ数nofURDの内容をデータ数信号nofDで置き換える。次にデータベース再構築制御部62は、学習数信号LNを読み出しデータ数信号RDNとして、データベース読み出し部61に送る。

【0188】データベース読み出し部61は、学習数信号LNを受けて、データベース記憶部60からLN個のデータを順次読み出し、必要な整形をして、情報フィルタリングユニット50に送る。

【0189】情報フィルタリングユニット50は、実施の形態1に記載した動作を行い未読データ記憶部10に格納する。

【0190】使用者は、インタフェースユニット51を起動し、未読データ記憶部10に格納されたLN個の未読データURDを順次読み出し、要不要を示す教師信号Tを入力する。LN個の未読データについて入力が終わると、使用者は学習開始信号入力端子106から学習開始信号LSを入力し、情報フィルタ装置の学習を行う。学習制御部指示信号出力端子107から出力される学習制御部指示信号LIが学習の終了を示すように1から0になると、データベース再構築制御部62は新たにLN個のデータを読み出すように読み出しデータ数信号RDNをデータベース読み出し制御部61に送り、新たにLN個のデータを情報フィルタリングユニット50を通し並べ変える。

【0191】使用者は、再び、インタフェースユニット51を起動し、LN個の未読データURDを必要か不要かを判断しながら、必要な情報が上位に来ているか否かを

確認し、情報フィルタ装置にさらに学習させるか否かを決める。

【0192】情報フィルタ装置の性能が不十分で、さらに学習させる場合には、使用者は、再び学習開始信号入力端子106から学習開始信号LSが入力し、情報フィルタ装置の学習を行う。

【0193】情報フィルタ装置の性能が十分に上がり学習が必要でなくなると、制御信号入力端子200からデータベース再構築実行を示す制御信号

$CDB=2$

を入力する。データベース再構築制御部62は、まず、スイッチ63を情報フィルタリングユニット50と適応データベース書き込み部64とが接続するように切り替える。次に、データベース再構築制御部62は、データベース記憶部60に記憶されたデータ数 $nofD$ 個のデータを読みだすように読み出しデータ数信号 RDN をデータベース読み出し部61に送る。データベース読み出し部61は、 $nofD$ 個のデータを順次読み出し、情報フィルタリングユニット50に送る。情報フィルタリングユニット50は、適応データベース書き込み部64の中のバッファにデータを必要性に基づいて並べ替える。

【0194】適応データベース書き込み部64は、書き込まれたデータの数が $nofD$ になると、バッファの内容を適応データベース記憶部65に書き込み、書き込み終了信号 EW をデータベース再構築制御部62に送る。書き込み終了信号 EW を受けたデータベース再構築制御部62は、データベース再構築制御指示信号出力端子210から出力されるデータベース再構築制御指示信号 IRD を1から0に変え、処理を終了する。

【0195】以上のようにデータベース再構築装置を構成することにより、使用者にとって必要な順にデータが並んだ使いやすい適応データベースを作ることができる。

【0196】なお、本実施の形態では、適応データベースは全体として元のデータベースと同じデータを持つようにしたが、記憶装置のメモリ領域を節約するために、適応データベースの内容をデータ間のリンク情報だけとしても、同じ効果が得られる。

【0197】（実施の形態4）以下、本発明の第4の実施の形態について、図面を参照しながら説明する。実施の形態4は、実施の形態2の情報フィルタ装置の構成にキーワード検索式生成部、キーワード評価信号ソート部及びキーワード評価部等を加えることにより、ユーザーに提示された「情報」について、必要／不要を解答するだけで、必要な情報を検索する検索式を自動的に生成することができるキーワード検索式生成装置を提供するものである。

【0198】図14にそのキーワード検索式生成装置のブロック結線図を示し、以下に説明する。

【0199】図14において、111はキーワード検索

式生成開始信号入力端子、112はキーワード検索式方法切り替え信号入力端子、113はキーワード検索式信号出力端子であり、30はメトリックの (i,j) 成分の重要性を評価するキーワード評価信号 $KWKD(i,j)$ を計算するキーワード評価部、31は前記キーワード評価信号を小さい順に並べ変えるキーワード評価信号ソート部、32は並べ変えられたキーワード評価信号により適応辞書信号を用いてキーワード検索式信号 E_q に変換するキーワード検索式生成部である。その他のブロックは、実施の形態2の情報フィルタ装置と同一構成であるので説明を省略する。

【0200】以上のように構成されたキーワード検索式生成装置の動作の前半のフローチャートを図15に、後半の動作のフローチャートを3つの方法に対応して図16、17、18それぞれに示す。以下、これらの図を参照しながら説明する。

【0201】動作の前半について、図15を参照しながら説明する。まず、キーワード検索式の生成を開始させるキーワード検索式生成開始信号 E_qGO がキーワード生成開始信号入力端子111から入力される。

【0202】キーワード検索式生成開始信号 E_qGO を受けたキーワード評価部30は、まず、回数記憶部25から全肯定回数信号 NY と全否定回数信号 NN とを、1次肯定メトリック記憶部26から1次肯定メトリック信号 $MY1$ を、1次否定メトリック信号記憶部27から1次否定メトリック信号 $MN1$ を読み出す（図15ステップ（イ））。

【0203】次に、キーワード評価部30は、今までに情報フィルタ装置が提示した情報についてユーザーが必要／不要の応答をしたかどうかを確認するために、全肯定信号 NY と全否定信号 NN の和 $(NY+NN)$ を計算し、さらに積 $NY \cdot NN \cdot (NY+NN)$ を計算する（同図ステップ（ロ））。前記積が0であることは、情報フィルタ装置はユーザーからどんな情報が必要であり、かつどんな情報が不要であるかを教えられていないことに対応するから、この場合、ユーザーの求めているキーワード検索式を推定することはできない。そこで、キーワード評価部30はキーワード検索式出力信号 E_q を

$E_q = (\text{キーワード検索式不明})$

としてキーワード検索式信号出力端子113から出力し処理を終了する（同図ステップ（ハ））。前記積 $NY \cdot NN \cdot (NY+NN)$ が0でない場合には、キーワード評価部30は、ユーザーが必要とした情報の割合と不要とした割合を示す情報出現割合 (Q_{yes}, Q_{no}) を $Q_{yes} = NY / (NY+NN)$
 $Q_{no} = NN / (NY+NN)$ と計算する（同図ステップ（ニ））。

【0204】確率的に考えて、ユーザーが必要とした情報に付いていた割合が前記 Q_{yes} より高いキーワード

は、ユーザーの必要とする情報を取り出す上で有効であると推定できる。実施の形態2で説明したように、前記1次肯定メトリック信号MY1と前記1次否定メトリック信号MN1には、対角成分にキーワード、非対角成分に2つのキーワードの共起についてのユーザーが必要／不要とした情報についての出現回数が記録されている

(以下、キーワードと2つのキーワードの共起をまとめてキーワードと表現する)。したがって、各成分についてQyes、Qnoと同様の割合の計算を行う。

【0205】このために、まずカウンタiを0にセットする(同図ステップ(ホ))。次にカウンタjを0にセットする(同図ステップ(ヘ))。

【0206】前記1次肯定メトリック信号MY(i, j)と前記1次否定メトリック信号MN1(i, j)の和(MY1(i, j)+MN1(i, j))を計算する。この和(MY1(i, j)+MN1(i, j))は、そのキーワードが過去何回生じているかを示す値である。この値が非常に小さい場合には、確率的にはあまり意味がないと考えられる。ここではこの和(MY1(i, j)+MN1(i, j))が3以上であれば、評価に採用するものとする(同図ステップ(ト))。この値(打ち切り値)は、3でなくとも別に構わないが、我々の実験では3から4程度が実用上便利であることが分かっている。

【0207】さて、この和(MY1(i, j)+MN1(i, j))が3以上であれば、キーワード出現割合(Pyes、Pno)を

$$Pyes = (MY(i, j) + \epsilon) / (MY(i, j) + MN(i, j) + 2\epsilon)$$

$$Pno = (MN(i, j) + \epsilon) / (MY(i, j) + MN(i, j) + 2\epsilon)$$

計算する(同図ステップ(チ))。ここで、 ϵ はPyesとPnoが0にならないようにするための0に近い正の定数である。

【0208】前記キーワード出現割合(Pyes、Pno)と前記情報出現割合(Qyes、Qno)の違いを表す量として、肯定偏差信号VY(i, j)と否定偏差信号VN(i, j)を

$$VY(i, j) = Qyes \cdot \log(Qyes / Pyes) \cdot \xi$$

$$VN(i, j) = Qno \cdot \log(Qno / Pno) \cdot \xi$$

ここで、

$\xi = \tanh[(MY(i, j) + MN(i, j)) / 3]$ である(同図ステップ(リ))。この係数 ξ は、出現頻度の多いものを比較的重要視するための工夫である。値3は、ここでは上で述べた打ち切り値と同じとした。必要に応じて、打ち切りより大きくとってもよい。こうして計算される肯定(否定)偏差信号は必要(不要)な情報に片寄って出現するキーワードについては、片寄りが大きいほど負の小さな値になるという性質を持っている。

【0209】和(MY1(i, j)+MN1(i, j))が2以下であれば、前記肯定偏差信号VY(i, j)と否定偏差信号VN(i, j)を

$$VY(i, j) = 0$$

$$VN(i, j) = 0$$

とする(同図ステップ(ヌ))。

【0210】以上の処理が終わると、カウンタjの値を1だけ増やす(同図ステップ(ル))。カウンタjの値が1次肯定／否定メトリック信号の行の数未満であると同様の処理を行い(同図ステップ(ヲ))、カウンタjの値が1次肯定・否定メトリック信号の行の数以上となると、カウンタiを一つだけ増やす(同図ステップ(ワ))。カウンタiの値が1次肯定／否定メトリック信号の行の数未満であると(同図ステップ(カ))、カウンタjの値を0にリセットし(同図ステップ(ヘ))、同様の処理を行い、カウンタiの値が1次肯定・否定メトリック信号の行の数以上となる処理を終了する。

【0211】こうして得られた肯定偏差信号VY(i, j)と否定偏差信号VN(i, j)とを、5つの値からなるキーワード評価信号

$$KWKD(\text{nofDiC} \cdot i + j) = (VY(i, j) + VN(i, j), VY(i, j), VN(i, j), i, j)$$

として出力する。最後の2つは、もともとの1次肯定／否定メトリックの成分がなんであるかを示すものであり、後にキーワードとの対応づけをするために必要なものである。

【0212】キーワード評価信号ソート部31は、キーワード検索式生成方法切り替え信号入力端子からのキーワード検索式生成方法切り替え信号MCKWに応じて、3つの方法でキーワード評価信号の並べ変えを行う(同図ステップ(ヨ))。

【0213】第1の方法(MCKW=1)は、必要な情報によく付けられるキーワードだけを取り出す方法であり、そのフローチャートを図16に示す。第2の方法(MCKW=2)は、不要な情報によく付けられるキーワードだけを取り出す方法であり、そのフローチャートを図17に示す。第3の方法(MCKW=3)は、両者を組み合わせた方法であり、そのフローチャートを図18に示す。以上、3つの方法を順番に説明する。

【0214】図16に示した第1の方法は、キーワード検索式生成方法切り替え信号MCKWが1の場合に選択されるものであり、キーワード評価信号KWKDを、その第2成分VY(i, j)について、小さい方から順に並べ変える(図16ステップ(イ))。このように並べ変えられたキーワード評価信号の第4、第5成分は、必要な情報に限って大きな値をとるキーワードを指し示す値である。そこで、キーワード信号ソート部は、この並べ変えられたキーワード評価信号SKWKDをキーワード検索式生成部32へと送る。

【0215】キーワード検索式生成方法切り替え信号MCKWが1の場合には、キーワード検索式生成部32は、適応符号辞書信号記憶部24から、ひとつの単位が文字列（キーワード）Wとそれに対応する符号Cと肯定回数PYと否定回数PNの4つからなる適応辞書信号FDCKをFDCK[1]からFDCK[nofFDCK]まで読みだす（同図ステップ（ロ））。

【0216】キーワード検索式の項の数を設定する方法には、手動による方法、自動で行う方法いずれも考えることができる。ここでは手動で行う方法を説明する。この場合、項数信号入力端子114から項数信号TNが入力される（同図ステップ（ハ））。

【0217】次に、キーワード検索式信号KWを0セットする（同図ステップ（ニ））。項数カウンタcountを1セットする（同図ステップ（ホ））。

【0218】並べ変えられたキーワード評価信号SKWKD(count)の第4成分iを読みだし、第i番目の適応符号辞書信号FDCK[i]の文字列を第1キーワードKW1へと変換、並べ変えられたキーワード評価信号SKWKD(count)の第5成分jを読みだし、適応符号辞書信号の第j番目の適応符号辞書信号FDCK[j]の文字列を第2キーワードKW2へと変換する（同図ステップ（ヘ）、（ト））。そして、キーワード検索式信号を

$$KW \leftarrow (KW) \text{ or } (KW1 \text{ and } KW2)$$

と置き換える（同図ステップ（チ））。

【0219】項数カウンタを1だけ増やす（同図ステップ（リ））。以下、同様の操作を、並べ変えられたキーワード評価信号SKWKD(2)、SKWKD(3)・・・SKWKD(TN)まで繰り返す（同図ステップ（ヌ））。並べ変えられたキーワード評価信号SKWKD(TN)まで上の処理を行いキーワード検索式信号KWを出力する。

【0220】なお、項数打ち切りを自動で行う方法の一つは、並べ変えに用いたキーワード評価信号が予め定められた値まで大きくなった時点で処理を打ち切る方法である。自動で行うもう一つの方法は、学習に用いた情報から必要な情報の全て（必要に応じて、90%、80%などに設定してもよい）を取り出せるようになるまで、処理を繰り返すという方法である。

【0221】図17に示した第2の方法は、キーワード検索式生成方法切り替え信号MCKWが2の場合に選択されるものであり、キーワード評価信号KWKDを、その第3成分VN(i, j)について、小さい方から順に並べ変えるものである。以下、前述した第1の方法と同様の処理を行い、orで結ばれたキーワード検索式KWを得る。しかし、MCKW=2の場合は不要な情報を取り出す検索式となっているから、最後にキーワード検索式信号をその否定

$$KW \leftarrow !KW$$

として処理を終わる（図17ステップ（ル））。

【0222】図18に示した第3の方法は、キーワード検索式生成方法切り替え信号MCKWが3の場合に選択されるものであり、キーワード評価信号KWKDを、その第1成分(VY(i, j) + VN(i, j))について、小さい方から順に並べ変えるものである。この方法は、肯定キーワード検索式信号YKWと否定キーワード検索式信号NKYとを中間表現として用いる方法である。

【0223】処理は、第1キーワード信号KW1と第2キーワード信号KW2を得るところまでは（図18ステップ（ト））、第1の方法と同様の処理であるが、以下の肯定偏差信号VY(i, j)の正負で切り替わる処理の部分から以降の異なる部分から説明する。

【0224】並べ変えられたキーワード評価信号SKWKD(1)の肯定偏差信号VY(i, j)が負である場合は、肯定キーワード検索式信号YKWを

$$YKW \leftarrow (YKW) \text{ or } (KW1 \text{ and } KW2)$$

と置き換える（同図ステップ（ヌ））。肯定偏差信号VY(i, j)が正である場合は、否定キーワード検索式信号NKWを

$$NKW \leftarrow (NKW) \text{ or } (KW1 \text{ and } KW2)$$

と置き換える（同図ステップ（リ））。この処理を、手動打ち切りの場合には、SKWKD(TN)まで行う（同図ステップ（ヲ））。

【0225】次に、キーワード検索式信号KWを

$$KW \leftarrow (YKW) \text{ and } ! (NKW)$$

と出力し（同図ステップ（ワ））、処理を終了する。

【0226】以上のようにして、本発明のキーワード検索式生成装置は、ユーザーが提示された情報について、必要／不要を解答するだけで、必要な情報を検索する検索式を自動的に生成することができる。

【0227】本発明の実施の形態4のキーワード検索式生成装置の実験による効果を図19に示し説明する。

【0228】図19に示した実験結果は、ユーザーによって必要／不要のラベル付けがなされた760件の情報をを用いた実験結果である。情報フィルタ装置に200件のデータを学習させ、残りの560件のデータを第1の方法を用いた本発明のキーワード検索式生成装置が生成したキーワード検索式によって検索を行ったときの検索効率を示している。横軸は、キーワード検索式の項の数、縦軸は検索された情報の割合を示している。実線は、ユーザーが必要とする情報の内何%が取り出されたかを、破線は全体の情報の内何%が取り出されたかを、点線は取り出された情報の何%がユーザーが必要とする情報かを示している。

【0229】以上、本発明によれば、項の数が約10程度で、必要な情報の約90%を取り出すことができ、その時取り出された情報のなかに含まれる必要な情報は約60%に昇っており、本発明が有効であることがはつき

り分かる。

【0230】

【発明の効果】以上のように、本発明は、情報に割り振られた複数のキーワードをベクトルに変換するベクトル生成部と、前記ベクトルと使用者からの教師信号を用いてスコアを計算するスコア計算部と、前記スコアから必要性和信頼性を計算する必要性計算部と、スコア計算部がスコアを計算する際に用いるメトリックを使用者から与えられる情報の必要／不要という単純な評価をもとに計算するメトリック学習部とを設け、情報をユーザーの必要度にしたがって並べ、ユーザーに対して必要性の高い情報から順に提供することにより、初心者にも精度の高い情報を得ることができ、更に使用者にとって必要性の高い情報の取り出し易い情報フィルタ装置を提供することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態1の情報フィルタ装置のブロック結線図

【図2】本発明の実施の形態1の情報フィルタ装置の概略を示すブロック結線図

【図3】本発明の実施の形態1の情報フィルタ装置のベクトル生成部の動作を説明するフローチャート

【図4】本発明の実施の形態1の情報フィルタ装置の未読データ書き込み制御部の動作を説明するフローチャート

【図5】本発明の実施の形態1の情報フィルタ装置の未読データ出力制御部の動作を説明するフローチャート

【図6】本発明の実施の形態1の情報フィルタ装置の学習制御部の動作を説明するフローチャート

【図7】本発明の実施の形態1の情報フィルタ装置のメトリック学習部の動作を説明するフローチャート

【図8】本発明の実施の形態1の情報フィルタ装置の判定面学習部の動作を説明するフローチャート

【図9】本発明の実施の形態1の情報フィルタ装置の判定面学習部の動作を説明するための図

【図10】本発明の実施の形態1の情報フィルタ装置の判定面学習部の動作を説明するための図

【図11】本発明の実施の形態2の情報フィルタ装置のブロック結線図

【図12】本発明の実施の形態2の情報フィルタ装置の辞書学習部の動作を説明するフローチャート

【図13】本発明の実施の形態3のデータベース再構築装置の概略を示すブロック結線図

【図14】本発明の実施の形態4のキーワード検索式生成装置のブロック結線図

【図15】本発明の実施の形態4のキーワード検索式生成装置のキーワード検索式生成の前半の動作を説明するフローチャート

【図16】本発明の実施の形態4のキーワード検索式生成装置の第1の方法を用いた場合のキーワード検索式生

成の後半の動作を説明するフローチャート

【図17】本発明の実施の形態4のキーワード検索式生成装置の第2の方法を用いた場合のキーワード検索式生成の後半の動作を説明するフローチャート

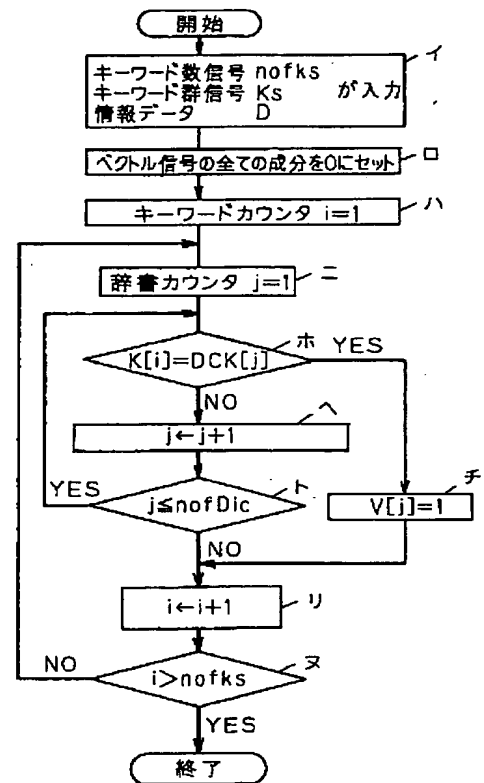
【図18】本発明の実施の形態4のキーワード検索式生成装置の第3の方法を用いた場合のキーワード検索式生成の後半の動作を説明するフローチャート

【図19】本発明の実施の形態4のキーワード検索式生成装置のキーワード検索式生成の効果の説明する図

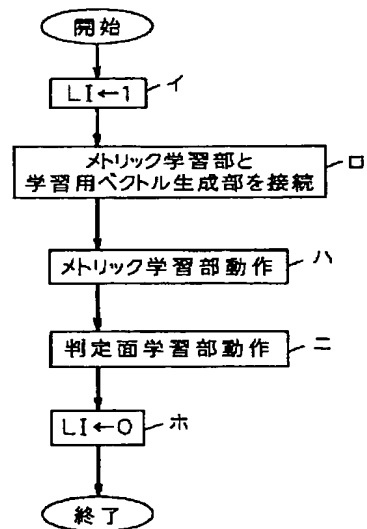
【符号の説明】

- 1 ベクトル生成部
- 2 辞書記憶部
- 3 スコア計算部
- 5 肯定メトリック記憶部
- 6 否定メトリック記憶部
- 7 必要性計算部
- 8 判定パラメータ記憶部
- 9 未読データ書き込み制御部
- 10 未読データ記憶部
- 11 未読データ出力制御部
- 12 教師データ制御部
- 13 教師データ記憶部
- 14 学習制御部
- 16 スイッチ
- 17 スイッチ
- 18 スイッチ
- 19 メトリック学習部
- 20 学習用ベクトル生成部
- 21 判定面学習部
- 22 スコア計算部
- 23 辞書学習部
- 24 適応符号辞書記憶部
- 25 回数記憶部
- 26 1次肯定メトリック記憶部
- 27 1次否定メトリック記憶部
- 28 KDメトリック学習部
- 30 キーワード評価部
- 31 キーワード評価信号ソート部
- 32 キーワード検索式生成部
- 50 情報フィルタリングユニット
- 51 インタフェースユニット
- 52 学習ユニット
- 60 データベース記憶部
- 61 データベース読み出し部
- 62 データベース再構築制御部
- 63 スイッチ
- 64 適応データベース書き込み部
- 65 適応データベース記憶部
- 100 情報入力端子
- 101 キーワード数信号入力端子

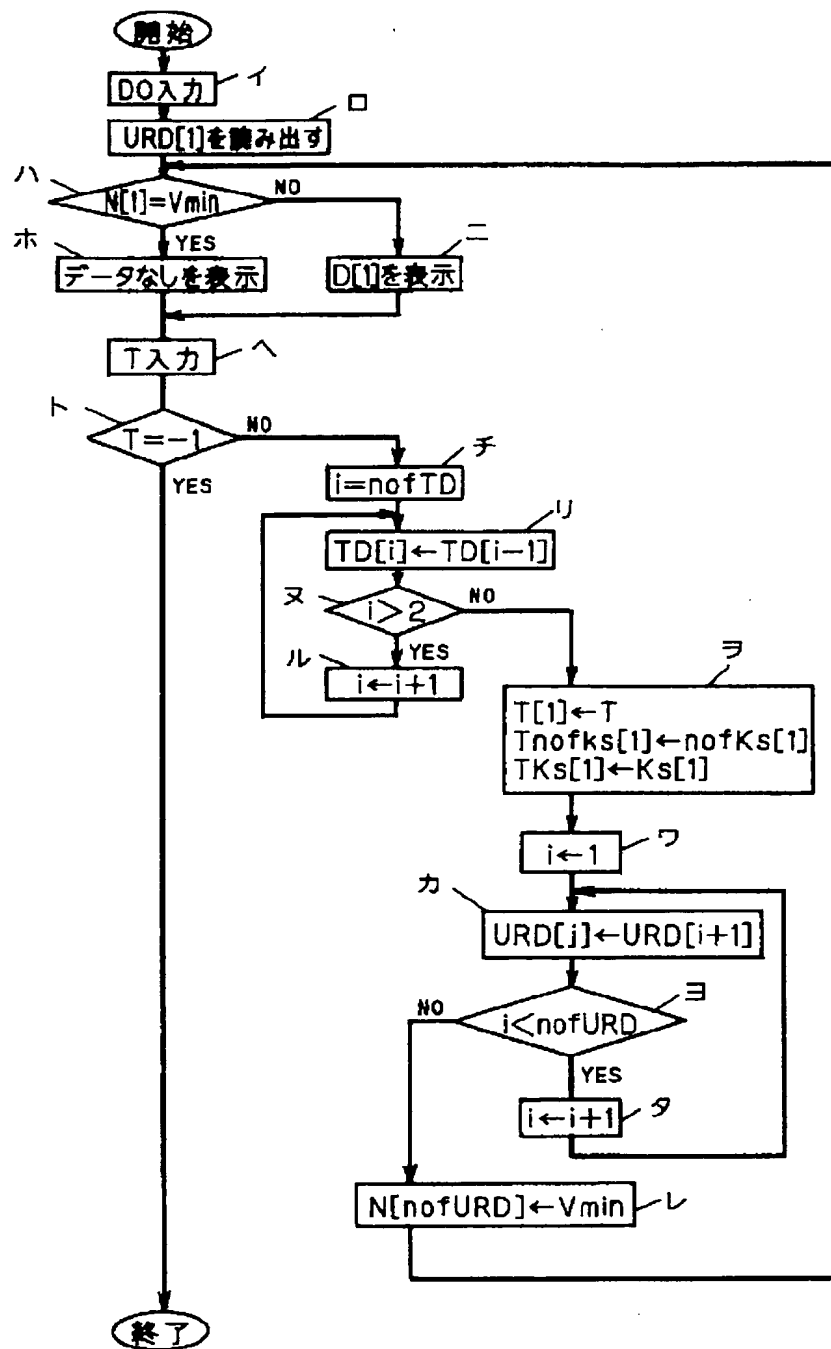
【図 3】



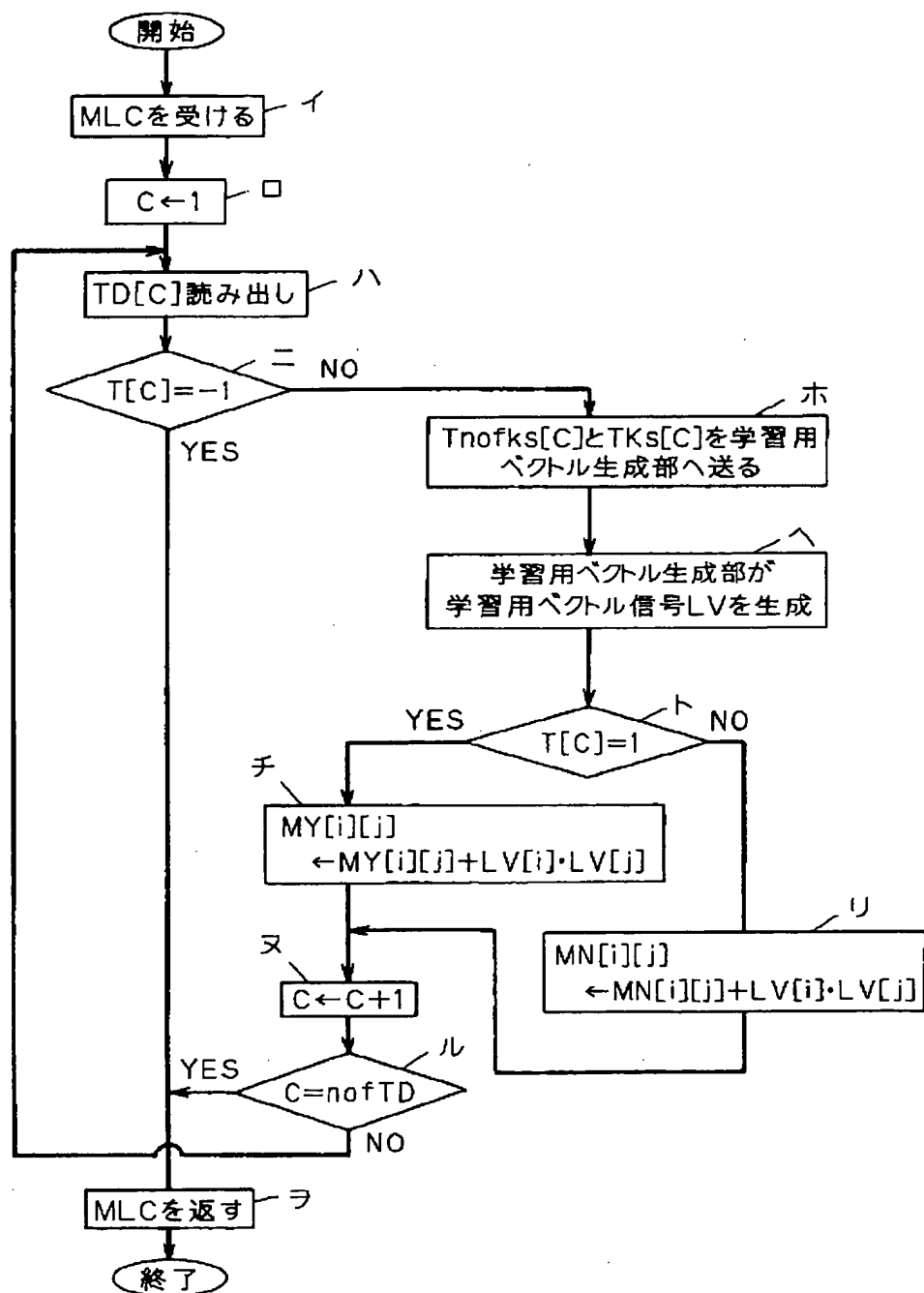
【図 6】



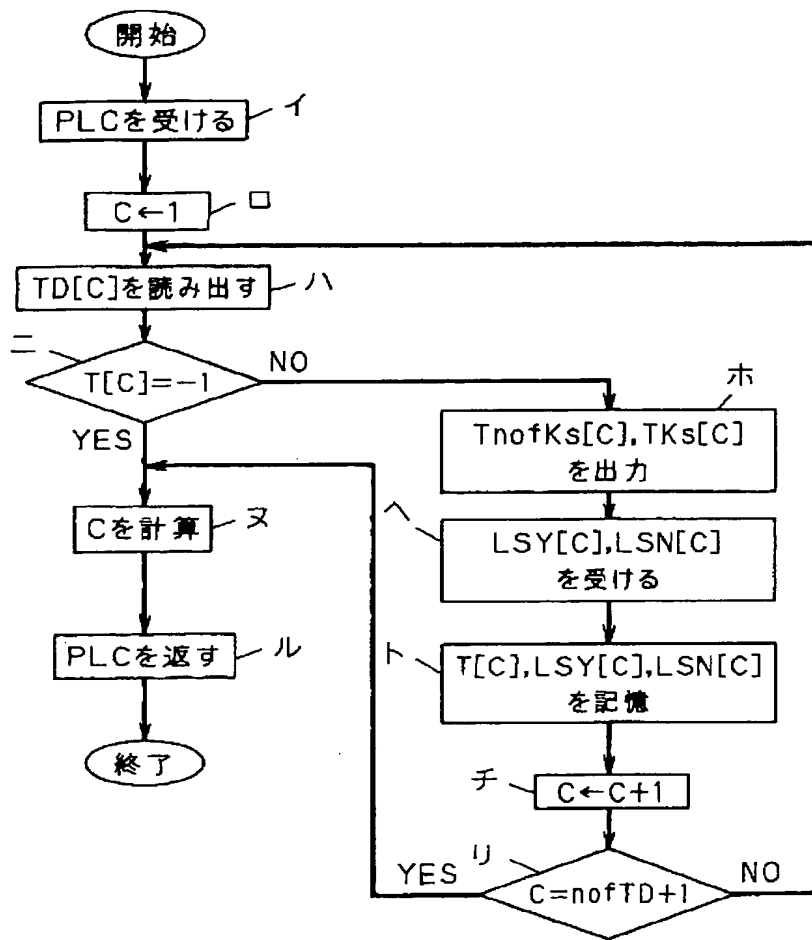
【図5】



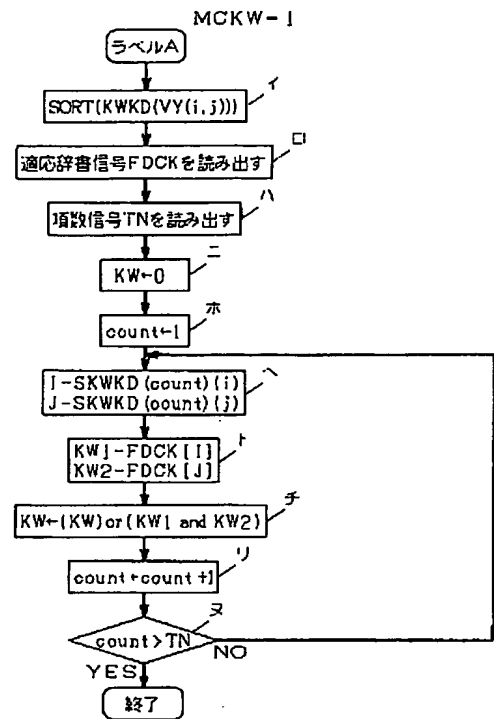
【図7】



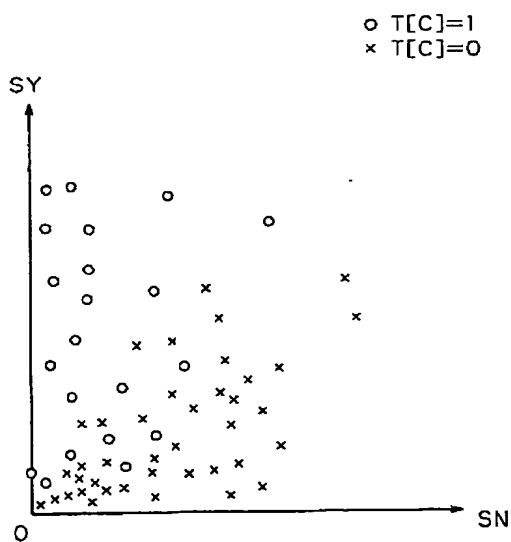
【図8】



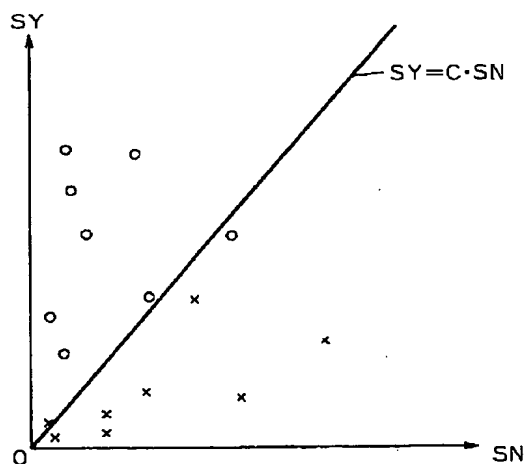
【図16】



【図9】

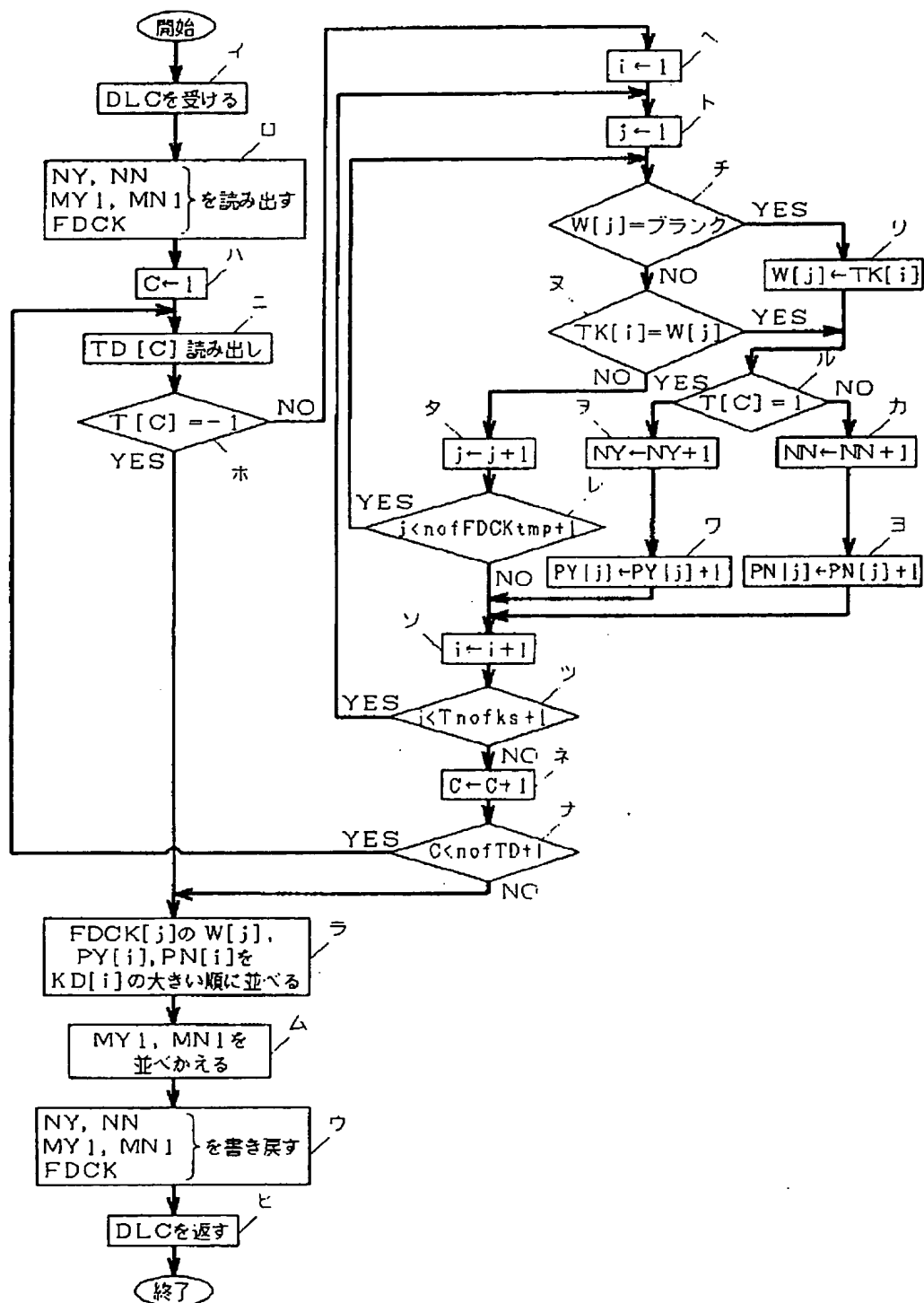


【図10】

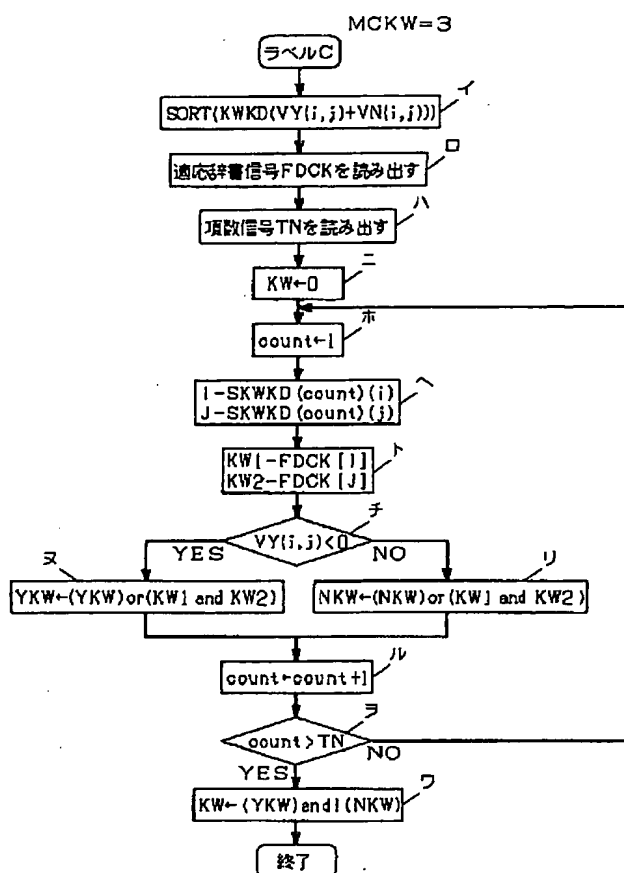
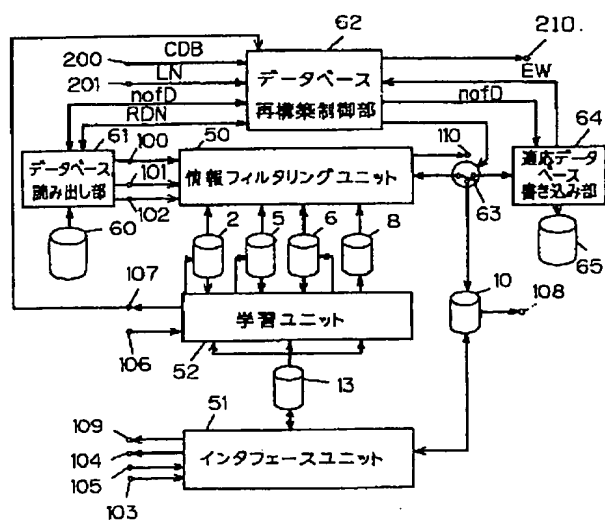


[illegible]

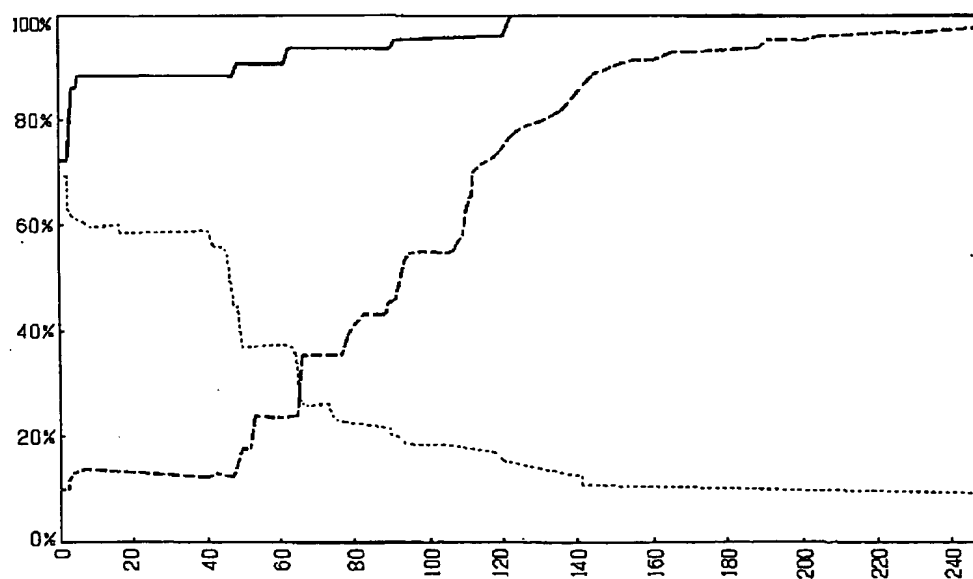
【図12】



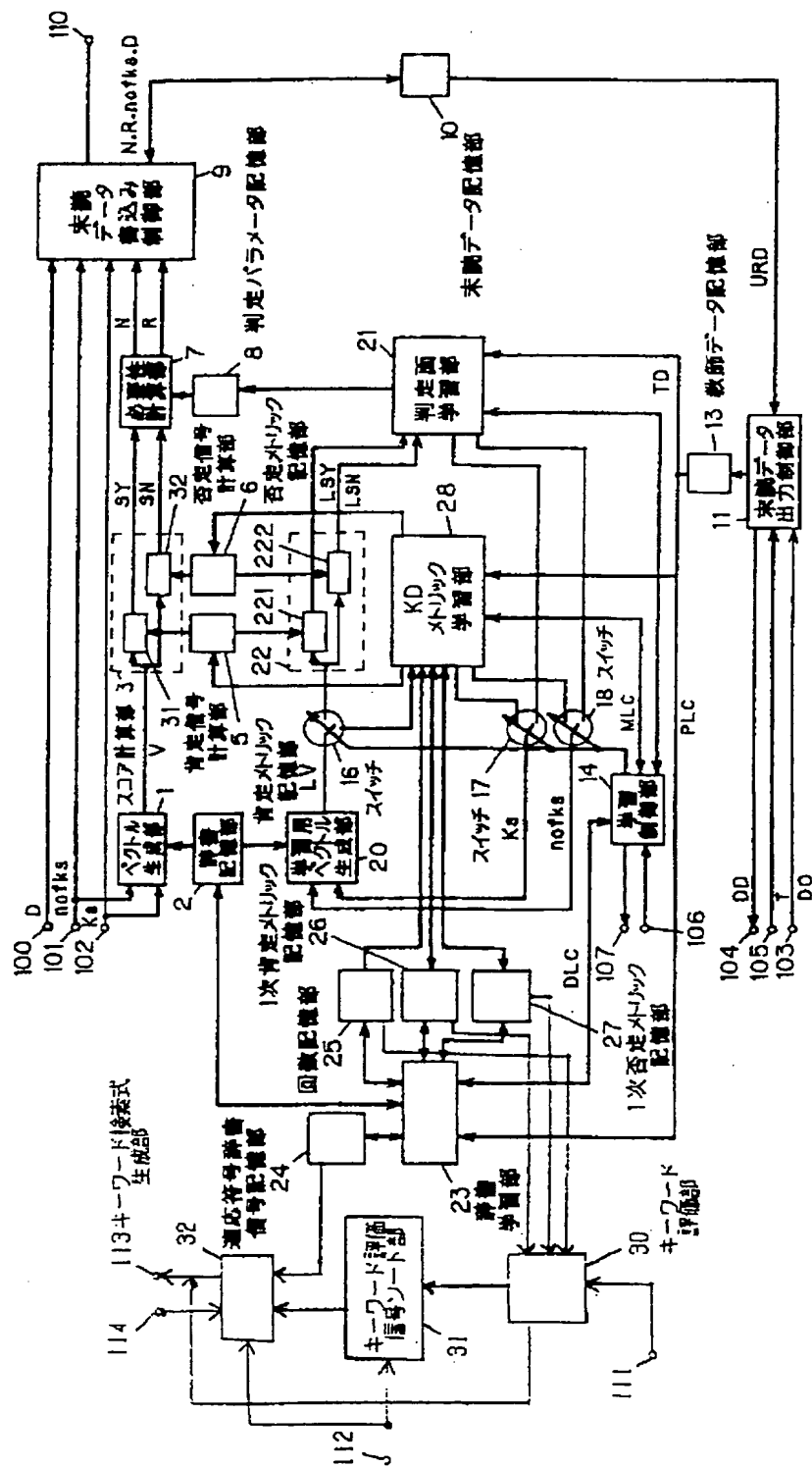
【图 18】



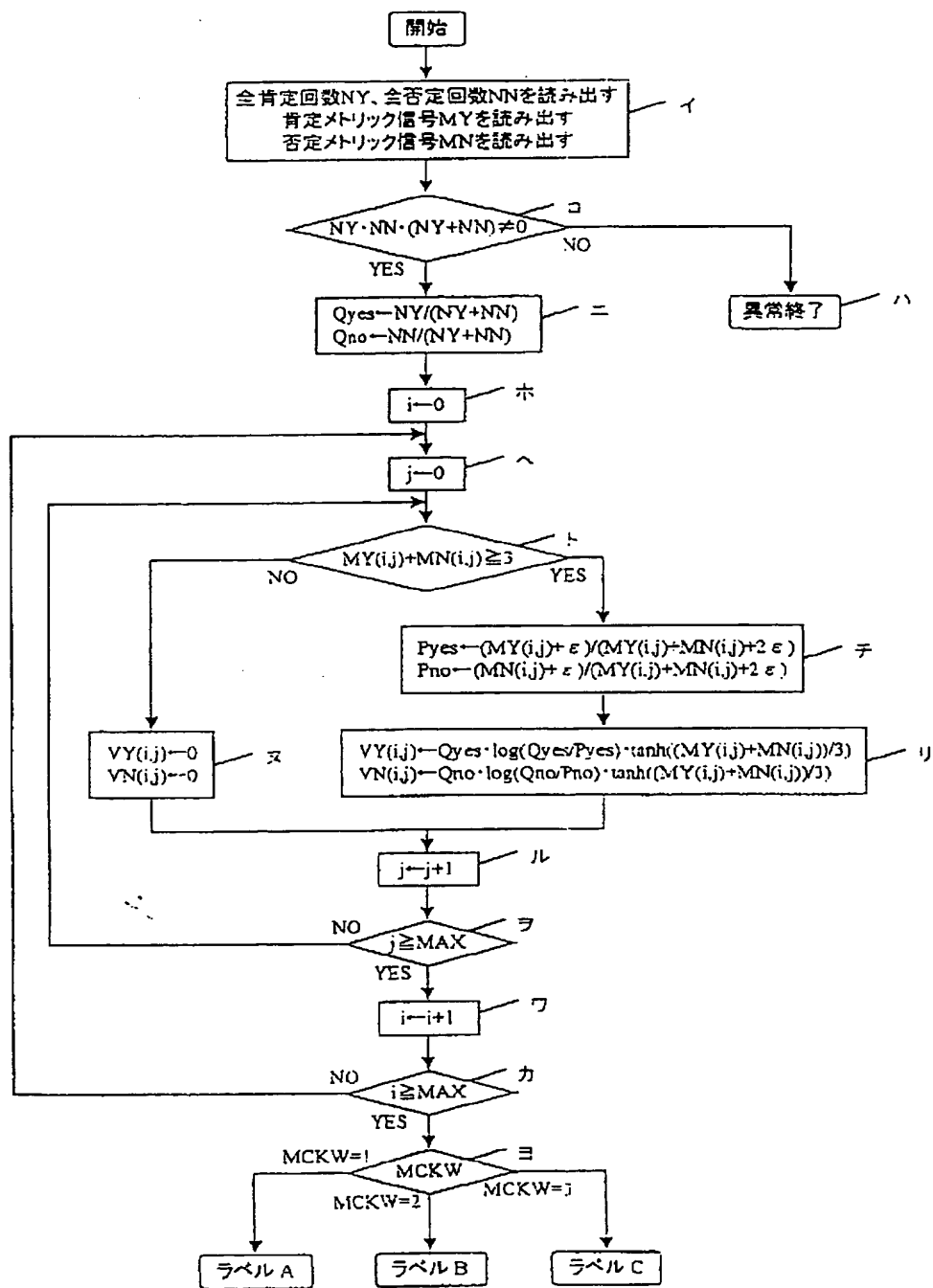
【図 19】



【图 14】



【図15】



【図17】

